"Math is a pain!": Understanding challenges and needs of the Machine Learning community on Stack Overflow

ZIHAN FANG, Vanderbilt University, USA
YU HUANG, Vanderbilt University, USA

Stack Overflow (SO) is a widely recognized online question-and-answer platform for programming, which has also fostered a substantial community dedicated to machine learning (ML), providing a space for both novices and experts to exchange ideas and find solutions to ML-related problems. However, as a relative minority of this online programming platform, research has demonstrated lower engagement in the ML community, but it remains largely unexplored to understand what hinders the engagement and contribution from ML users' perspectives. This paper presents an empirical study based on 22 hours of semi-structured interviews and 131 survey responses with users on SO and reveals the key factors that may lead to the lower response rate and extended waiting time for ML questions on SO, which includes the unique quality requirement for posting ML questions, the discrepancy between time invested and benefits gained, the dispersed nature of the ML community across various platforms and the desired improvement for SO. Moreover, the qualitative study reveals a declining friendliness in SO's culture over time; the subsequent quantitative study corroborates that newcomers frequently encounter stress when posting and answering ML questions, even though this stress diminishes with increased experience. Additionally, we also explored the potential influence of generative AI tools (e.g., ChatGPT) on online question-and-answer platforms, specifically focusing on ML Q&A. We hope the results of this study can pave the way for enhancing the experience of ML users on online platforms, ultimately facilitating improved knowledge exchange and collaboration within the ML domain.

CCS Concepts: • **Human-centered computing → User studies**.

Additional Key Words and Phrases: Machine Learning, Stack Overflow, ChatGPT

ACM Reference Format:
Zihan Fang and Yu Huang. 2024. "Math is a pain!": Understanding challenges and needs of the Machine Learning community on Stack Overflow. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 451 (November 2024), 36 pages. https://doi.org/10.1145/3686990

1 Introduction

The field of machine learning (ML) is experiencing rapid growth and high demand [1], attracting both experienced developers and newcomers from diverse domains. ML is a specialized field of artificial intelligence (AI) that emphasizes the enhancement and refinement of analytical capabilities through computational algorithms [2], encompassing a variety of subfields, such as supervised learning, unsupervised learning, reinforcement learning, deep learning, natural language processing (NLP), computer vision, and more [3]. Different from traditional programming, ML implementation requires both programming skills and mathematical principles [4]. Essentially, ML programming entails designing algorithms for tasks such as data preprocessing, model creation, evaluation, and deployment [5]. Concurrently, ML also draws upon mathematical principles such as statistics,

Authors' Contact Information: Zihan Fang, zihan.fang@vanderbilt.edu, Vanderbilt University, USA; Yu Huang, yu.huang@vanderbilt.edu, Vanderbilt University, USA.

ACM 2573-0142/2024/11-ART451
https://doi.org/10.1145/3686990

linear algebra, and calculus to offer theoretical understanding and refine methodologies [6]. As ML continues to mature, its applications are becoming increasingly widespread. Various industries are using ML to solve technical issues [7] and improve work efficiency [8]. Meanwhile, Stack Overflow (SO), a widely renowned question-and-answer platform [9], has emerged as a crucial resource for addressing ML challenges, especially those related to model building [10]. Recent research has revealed that ML questions on SO are answered less frequently and have an average of ten-time longer waiting period to be answered than other general programming questions [11], which may lead to frustration and a lack of confidence among developers seeking assistance, hindering their ability to learn and apply ML techniques effectively. Moreover, this phenomenon may negatively affect the overall growth and sustainability of the ML industry, potentially discouraging developers from continuing to pursue or contribute to the field. In the past, to improve the participation and experience on online Q&A platforms (i.e., SO) for general programming-related questions, researchers have proposed several approaches, such as methods for recommending related questions to developers based on semantic matching [12], editing tools to identify minor text problems in posts and recommend sentence edits for correction [13] and transformer-based post title generation approaches to improve the quality of questions on SO [14]. Some researchers have explored the questions that software developers always ask when learning about and using ML libraries on SO [15]. However, these existing studies primarily focus on the general SO community, whereas the domain of ML, being an interdisciplinary field, exhibits distinct patterns of question-and-answer behavior [15, 16]. In particular, ML questions frequently center on data comprehension and modeling, demanding a combination of programming skills and domain-specific knowledge, such as an advanced mathematical concept and specialized algorithmic understanding [17]. The general programming questions usually deal with coding problems or explaining programming concepts and basics [18]. Thus, a broader community on SO, including software developers, computer scientists, and programming enthusiasts, focuses more on general programming problems [19]. On the contrary, ML questions are more likely to be contributed by experts in ML, mathematics, or data science, who are not the majority on this platform [11]. However, as the importance and complexity of ML continue to grow in the contemporary technological landscape [20], the ML community is still facing challenges including more complex questions, decreased engagement, and extended response times [11] that may pose significant obstacles to fostering a vibrant, inclusive ecosystem that leverages collective expertise and accelerates advancements. Despite the recognition of these challenges, the underlying reasons for these phenomena remain largely elusive. To address the low engagement and improve the ML community accordingly, a comprehensive investigation is imperative to identify the specific needs and challenges of developers within the ML community, as well as to explore the difference in their interactions compared to other programming communities on SO or other similar forums.

In this study, we first conducted semi-structured interviews with 24 users who have been active under ML Q&A on SO to identify the common needs and challenges faced in the ML community when browsing, posting, and answering questions. Subsequently, we applied thematic analysis to the qualitative interview data to identify key themes and patterns. To further investigate the interaction patterns and challenges within the ML community on SO, based on the qualitative analysis results from the interviews, we then designed and distributed a survey and collected 131 responses from SO users to (1) verify the needs and challenges identified in the thematic analysis and (2) examine the disparity in user engagement between the ML community and general programming community on SO. Importantly, our study offered a novel comparison of the interactions of the ML community with those of other programming communities on SO and similar platforms from users' perspectives, a focus not explored in previous research. Additionally, due to the occurrence and popularity of ChatGPT during our study period, we also explored how ChatGPT influenced the

developers' usage of online Q&A platforms in our study and proposed possible approaches using large-scale language models to improve the ML community.

Subsequently, we found that users typically need to browse more Q&A threads on SO to resolve their ML questions compared to general programming questions, and ML-related questions are found to be dispersed across multiple platforms instead of being concentrated on a single platform. Formulating ML questions and answers is also constrained by time, which acts as a hindrance to users' motivation to actively participate. Furthermore, the ML community of SO exerts pressure on ML users, particularly new users. This pressure tends to decrease as users gain more experience, but they still experience higher levels of stress compared to engaging in general programming questions. During our study, the emergence of ChatGPT has resulted in a decline in the frequency of SO usage, despite users displaying less trust in the answers provided by ChatGPT when compared to SO. The main contributions of this study are the following:

- Investigated the disparity in user engagement between the ML community and general programming community on SO.
- Identified the specific challenges faced by developers within the ML community on SO as well as the potential factors that lead to the challenges.
- Uncovered the impact of the generative AI models (i.e., ChatGPT) on question-and-answer platforms and proposed potential opportunities from the developers' perspectives.
- Provided suggestions to improve collaboration and user engagement within the ML community.

With ML being a widely applicable technology, we believe it is important to effectively stimulate the progress of ML and support the community with a full understanding of its unique characteristics and challenges. Ultimately, our primary objective is to learn the factors leading to the challenges confronted by developers when utilizing online Q&A platforms for ML problems and guide the design of approaches to enhance user engagement and interaction within the community.

2 Related Work

2.1 Challenges in Learning Machine Learning

Numerous fields have witnessed significant advancements in ML, such as medicine [21], materials science [22], pharmacometrics [23], geoscience [24], and education [25]. As a result, ML has transcended from being an area of academic research interest to a practical method that can address real-world challenges [26], facilitating growth in various industries beyond its domain [27]. However, for individuals, especially novices in the field, acquiring knowledge and skills in ML poses several challenges [28]. These challenges include limited access to learning materials, lack of comprehension regarding the complexities of ML systems, overwhelming coursework demands, inadequate instructor preparation, and technological limitations associated with the current educational methodology utilized for ML [29]. In addition, ML, being an inherently interdisciplinary field [30], necessitates individuals to acquire knowledge from diverse domains. The different programming languages with different libraries and frameworks also increase the challenges in learning and implementing ML [31]. Faced with the above issues, engaging in the ML community demonstrates advantageous implications for acquiring knowledge in this field [32], where learners are afforded the opportunity to receive help from other users [33], thereby facilitating a rapid and comprehensive understanding of ML concepts and implementation. Our investigation takes a mixed method of quantitative and qualitative study [34] to comprehend the difficulties involved in ML knowledge sharing within the context of the online ML community.

2.2 Challenges in Using Stack Overflow

SO has emerged as a popular online platform utilized by developers for the purpose of acquiring and exchanging knowledge [35]. However, users may encounter challenges when engaging in this platform in diverse contexts, such as academic research [36] or industrial work [37]. Due to its public nature, users voluntarily contribute their knowledge and seek assistance [38], which renders the Q&A on SO susceptible to issues such as the presence of duplicate questions [39], provision of incomplete or inaccurate answers [40], and the prevalence of outdated information [41]. Regrettably, there is currently no effective approach has been employed by the platform to address these concerns though many approaches have been proposed in academia [12, 42, 43]. Furthermore, although SO predominantly caters to an audience focused on programming-related matters [44], it still attracts numerous users who engage in discussions pertaining to various other subjects, such as mathematics [45], data science [46], and so forth. However, it is worth noting that these questions related to unconventional topics experience lower levels of activity in comparison to general programming questions [11].

2.3 Machine Learning Questions on Stack Overflow

SO has also become the preferred platform for people to ask ML questions [47] due to the absence of an integrated platform for this purpose. A study conducted by Islam et al. explores the problems people encounter when learning and using ML and asking questions on SO [15]. Another study investigates changes in ML-related questions on SO over time and across six programming languages to better understand the ML challenges in different programming languages [48]. Additionally, Alshangiti et al. analyze ML-related posts on SO to investigate the challenges of developing ML applications, finding that ML questions on SO have a lower response rate and fewer ML experts compared to general programming domains [11]. In particular, Alshangiti et al. highlight that ML questions on SO take ten times longer to receive answers than general programming questions. They also find that ML questions requiring both conceptual understanding and implementation expertise often lack accepted answers, leading them to conclude that ML questions are more challenging to answer on SO. Researchers have attempted various methods on SO to improve the situation. For example, multiple practical tools have been developed to help improve the quality of questions on SO, including detecting duplicate questions with deep learning approaches [42], predicting the quality of questions with NLP approach [43], and recommending related questions to developers based on semantic matching [12], etc. However, although previous approaches have aimed to improve the quality of questions from a broad perspective, there is a necessity for more focused investigation when it comes to ML questions that display interdisciplinary characteristics [30]. Previous research (i.e., [11]) has identified certain issues within the ML community on SO, but it did not delve into the underlying causes of these problems. In this study, we conducted an in-depth analysis of the potential reasons behind these challenges, aiming to offer more focused insights and develop specific strategies to address them.

3 Study Design and Methodology

To investigate how developers address ML questions and provide answers on SO, as well as to identify the challenges they encounter on the platform, we conducted a comprehensive study employing a mixed-methods approach [34]. This involved (1) conducting semi-structured interviews with users who have actively engaged in the ML community on SO, and (2) based on the knowledge learned from the interviews, designing and distributing a survey to a diverse range of SO users with a specific interest in ML. Our study was conducted under the protocol reviewed and approved by our local Institutional Review Board (IRB).

Specifically, in this study, we aimed to investigate the following research questions:

- RQ1: How do developers on SO search, ask, and answer ML questions?
- RQ2: What challenges are developers facing when using SO for the purpose of ML Q&A?
- RQ3: How can we improve and support the ML community on SO?
- RQ4: What is the impact of the emergence of generative AI tools (i.e., ChatGPT) on SO specifically for the ML community?

For the replication purpose of the study, the interview script, survey design, and codebook used in the qualitative analysis are included in the supplementary package and will be released publicly after the review process.

3.1 Semi-Structured Interviews

Semi-structured interviews are already commonly applied in diverse studies as they provide the flexibility to explore more tailored questions based on the personal experiences of interviewees, especially for topics that are not yet well explored [49, 50]. Thus, to understand the experiences and challenges of developers in the ML community on SO with little previous knowledge, we first designed and conducted semi-structured interviews with ML users on SO and then applied qualitative analysis to the interviews.

Participants: A total of 24 developers were recruited remotely as interview participants based on their involvement in ML Q&A activities on SO. The selection process involved purposely selecting the top 1,000 users with the highest number of accepted answers for ML-related questions. This selection was performed using Query Stack Exchange, an open-source tool that allows for executing custom queries on public data from SO.[1] To ensure the participants' relevance to the ML community, the first author manually double-checked their engagement in ML topics and emailed the users who have valid activity history under ML topics (e.g., machine learning, NLP, artificial intelligence, time series, regression, linear regression, transformer, computer vision, deep learning, unsupervised learning, supervised learning, reinforcement learning, neural network, TensorFlow, Keras, PyTorch, scikit-learn, adopted from [11]). During the recruitment process, we adhered strictly to our approved IRB. When gathering email addresses from potential participants, we only obtained emails from users who explicitly indicated their willingness to be contacted on their public websites. Initially, we sent these individuals an email providing a brief overview of the study along with a consent form. Once they had signed the consent form, we proceeded to send them subsequent emails outlining the specifics of the study and scheduling interviews. In total, we contacted 218 SO users via email, resulting in a response rate of 14.7% (i.e., 32 out of 218). Eventually, 24 meetings were scheduled with developers who agreed to participate online. Table 1 presents a summary of the demographic information of the interview participants, including their gender, country of origin, ML experience, and their usage of SO. Additionally, information regarding their educational background (e.g., major) and programming experience was collected and can be accessed online. Among the 24 participants, all individuals possess over three years of programming experience, while five participants have less than three years of experience specifically in ML. Among these five participants, their exposure to SO is also less than three years, with only one user having more than three years of experience using the platform. The remaining participants have more than three years of experience in both learning ML and using SO.

Interview Questions Design: The semi-structured interview comprises 29 questions, labeled I1[2] through I29, and explores participants' experiences with ML Q&A on SO in three key aspects:

[1] https://data.stackexchange.com/stackoverflow/queries

[2] I1 represents the first interview question, a format maintained across all subsequent interview questions. For example, I1 through I29 denote the entire set of 29 interview questions.

browsing ML Q&A (I1-I5), posting ML questions (I6-I13), and answering ML questions (I14-I24). The interview concluded with questions about SO features (I25-I27), reasons for unanswered or delayed ML responses (I28), and suggestions for enhancing the ML community on SO (I29). Regarding browsing behavior, we start with questions about how they use SO for ML-related questions encompassing code reuse, voting behaviors, and engagement with various answers (I1). Further questions explore search patterns, keyword use, and collection of multiple answers (I2, I3). Participants also reflect on challenges such as language barriers, discrepancies in solution tasks (I4), and negative experiences on SO or similar platforms (I5). In terms of posting behavior, participants share the types of ML questions they commonly post (I6-I9) and their expectations for response times and answer acceptance rates (I10, I11). The interview also explores perceptions of community inclusivity, conflicts, and peer influences (I12, I13). The remaining interview questions are designed to explore participants' answering behavior starting with exploring participants' motivations, selection criteria, and engagement in answering ML questions (I14, I15, I17-I19). Subsequently, we examine the challenges faced and negative experiences on the platform (I16, I19, I20, I22, I23). Additionally, participants reflect on the significance of their involvement and emotional factors like altruism or self-realization associated with answering ML questions on SO (I21, I24).

Thus, answers to RQ1 stem from the questions related to browsing behaviors (I1-I3), posting behaviors (I6-I9), and answering behaviors (I14-I15 and I17-I19), totaling 12 questions. To tackle RQ2, we have designed a series of 13 questions that encompass challenges encountered during browsing (I4-I5), posting (I10-I13), and answering (I16, I20-I24, and I28). Furthermore, insights for RQ3 can be garnered from questions I25-I27 and I29.

Protocol: We first conducted a pre-interview survey to learn about participants' demographic information, educational and professional backgrounds, as well as their experience in the domains of ML and SO. Then, the first author conducted remote semi-structured interviews with a duration ranging from 30 to 60 minutes, with the variation in time being attributed to the participant's level of engagement on SO. For instance, individuals who only post ML questions (i.e., they never answered questions on SO) tend to have shorter interviews. As a token of appreciation for their participation, each participant received a compensation of $35 USD in the form of an electronic Amazon gift card upon completion of the interview. All interviews were recorded and subsequently transcribed using a transcription tool (Otter.ai[3]) due to its cost-effectiveness, speed, and an approximate 80% match to manual transcription in general [51]. However, it may fail to capture the interactional intricacies inherent in human dialogues, potentially leading to the omission or mistranscription of interactional data [51]. To eliminate the limitation, the first author thoroughly reviewed three transcripts against their corresponding original audio recordings to ensure content consistency. Additionally, during thematic analysis, the authors also revisited the audio recordings to clarify ambiguous textual elements. The first three interviews were conducted as pilot interviews, incorporating a comprehensive range of questions, which were subsequently refined to form the final interview script for the remaining interviews. During each semi-structured interview, we may slightly modify certain questions or come up with the following questions based on the participants' responses obtained from the pre-interview survey. Each interview was conducted in English.

Analysis: In total, we collected a total of 22 hours of semi-structured interviews from 24 participants. We employed the thematic analysis method [52] on the interview transcripts using ATLAS.ti[4]. Drawing from the interview script, we initially identified five distinct categories and proceeded to generate open codes by identifying significant recurring statements made by the

[3]https://otter.ai
[4]https://web.atlasti.com

participants. To ensure reliability, the first author independently analyzed three transcripts to identify open codes, followed by discussions and code relationship determinations with other authors. The first author then iteratively and collaboratively refined the labeling process based on the refined code book. Subsequently, the authors worked together to identify relationships among the codes, resulting in the organization of the codes into 27 coherent and meaningful themes. Theoretical saturation, indicating the point at which no new information or insights are gained from further interviews [53], was reached at P21. Subsequent participants, P22 through P24, did not contribute further unique information. For reference, our code book is accessible in the supplementary package.

3.2 Survey

To validate our themes and findings derived from the thematic analysis of the interviews and further explore the research questions, we designed and conducted a survey targeting a more expansive cohort of developers within the ML community on SO. Our recruitment approach mirrored that of the interviews, and we identified the top 7,000 users who possessed the greatest number of accepted answers on ML-related questions using Query Stack Exchange, from which we identified a subset of 1,830 users for whom contact information was publicly accessible on SO. We followed the same recruitment process as our qualitative study.

Survey Questions Design: In the interview, numerous participants express they are more active in the general programming community on SO, while some participants mention a decrease in their usage frequency of SO as a result of the emergence of ChatGPT. Thus, our survey is designed to include three main themes of questions, comprising 64 questions labeled S1[5] through S64. The survey aims to: (1) gather respondents' experiences within the ML community on SO and other similar platforms, (2) determine if respondents have contributed to the general programming community before and the disparities they observe compared to the ML community, and (3) gather their opinions on ChatGPT in comparison to SO for addressing ML problems, thereby providing further validation for our qualitative findings. The survey includes multiple-choice questions, Likert-scale items (such as a 5-point scale ranging from "Strongly Disagree" to "Strongly Agree"), matrix tables where participants evaluate statements using a 5-point Likert scale, and open-ended questions.

Specifically, the survey encompasses 15 questions collecting demographic data (S1-S15). For these questions, we adopted the race and ethnicity descriptions from the certified questions[6] provided by Qualtrics[7], while the gender descriptions followed the guidelines established for HCI researchers [54]. Among all the other survey questions, there are 11 questions (S16-S26) about browsing behavior, 16 questions (S27-S42) focused on posting behavior, and 16 questions (S43-S58) concerning answering behavior. Within each behavioral section, we incorporated participants' comparative experiences with ML and general programming Q&A. To ensure accurate responses, we highlighted key text in these questions. Furthermore, due to several interviewees expressing a preference for ChatGPT, we included two questions (S59-S60) inquiring about their views on using ChatGPT for ML problems. Concluding the survey, three questions (S62-S64) evaluate participants' experiences with SO, including their attitudes, expectations about its features, and general improvement suggestions. Regarding browsing behavior, participants assess aspects such as code reuse, voting, post-viewing, and search methods for ML Q&A (S16-S17, S20-S24). They

[5]S1 represents the first survey question, a format maintained across all subsequent survey questions. For example, S1 through S64 denotes the entire set of 64 survey questions.

[6]https://www.qualtrics.com/support/survey-platform/survey-module/editing-questions/question-types-guide/pre-made-qualtrics-library-questions/

[7]https://qualtrics.com/

are also requested to contrast SO's community culture and inclusivity for ML Q&A with general programming (S18-S19, S25-S26). In terms of posting behavior, participants elaborate on the types of ML questions they commonly post, techniques for phrasing questions, their confidence in formulating effective questions (S27-S28, S31-S33, S37), and other similar forums they prefer (S29-S30). The survey also evaluates community inclusivity and responsiveness, comparing ML and general programming on SO, considering stress, expectations, and response times (S34-S42). Addressing answering behavior, the survey inquires about participants' motivations, search habits, challenges, preferred similar platforms (S43-S48), and perceptions of community attitudes (S49-S55). Then, the survey contrasts ML and general programming on SO regarding stress, expectations, and confidence (S56-S58). Additionally, in line with the interviews, we also collect opinions regarding the issue of low response rates within the ML community on SO (S61). The full design of the survey is included in the supplementary package.

Thus, RQ1 can be addressed using data from S16-S17, S20-S24 (browsing behavior), S27-S33, S37 (posting behavior), and S43-S48 (answering behavior). For RQ2, we consulted S18-S19, S25-S26 (browsing), S34-S36, S38-S42 (posting), S49-S58 (answering), and S61. S62-S64 can address RQ3, and S59-S60 pertains to RQ4.

Protocol: The consent form is situated on the first page of the survey instrument. Participants proceed to the survey questions only after providing their agreement and affixing their signature on the consent form. The respondents were able to complete the survey in approximately 20 minutes. The majority of the survey questions were designated as optional. To minimize the likelihood of participants misinterpreting the questions, the survey underwent a thorough review by three individuals who were neither co-authors of this paper nor potential participants in the study. The pilot survey was conducted electronically using Qualtrics[8], after which participants were emailed to gather feedback. The survey used for pilot tests consisted of 64 items mirroring the survey employed in the subsequent quantitative study. In light of the pilot test feedback, modifications were made to enhance the survey's reliability. These revisions involved simplifying language, clarifying ambiguous questions, and incorporating additional context as needed to ensure consistent interpretation of questions among respondents. Subsequently, the survey was piloted with a small group of 20 individuals before being distributed on a larger scale. During the second pilot phase, we used the modified survey from the first pilot phase and incorporated an additional question to gather feedback on the survey's content and structure. Specifically, participants were requested to provide feedback regarding the relevance of the questions and any other issues they encountered during their participation or overall experience with the survey. Upon integrating the feedback and implementing requisite modifications, we distributed the finalized survey to a broader cohort of potential participants, excluding the feedback question. Following completion of the survey, participants were given the opportunity to enter sweepstakes for a chance to win one of five electronic Amazon gift cards valued at $50 USD each.

Analysis: Out of the 1,830 email invitations sent, we received 149 responses with 131 being valid (i.e., the survey is completed), resulting in a valid response rate of 7.16%, which is consistent with previous survey studies on online platforms for programming [55]. Specifically, our respondents consisted of 120 males and 8 females (3 participants who preferred not to disclose their gender). They came from various regions: Africa (n = 4), the Americas (n = 41), Asia (n = 34), Europe (n = 51), and Oceania (n = 1). Furthermore, the respondents exhibited diverse levels of experience in both ML and SO, although all of them possessed programming experience exceeding one year. Among the valid responses, 111 users reported previous engagement with browsing ML questions and answers, 98 had posted ML questions, and 74 had answered ML questions on SO.

[8]https://www.qualtrics.com/

Moreover, we employed various statistical tests to evaluate the significance of observed differences in the quantitative results [56]. Initial analysis was conducted using descriptive statistics to provide an overview of the data, helping to set the stage for further statistical tests [57]. We applied the z-test to evaluate percentage score differences for statistical significance [58], determining whether these variations signify discrepancies within the broader population [59]. Moreover, we used the t-test to distinguish between the means of two groups or paired measurements [60], which can examine participants' engagement with general programming versus ML questions on SO, including aspects like time allocation and usage patterns across SO features. Additionally, the chi-square test was employed to validate variations in categorical variables [61] due to the questions with categorical responses, such as perceptions of SO culture. Finally, considering the limited sample size, the effect size was incorporated to elucidate the practical relevance of our findings [62].

ID	Gender	Age	Country	ML Exp	SO Exp
P1	Man	25-34	United States	> 3 years	> 3 years
P2	Man	35-44	United States	> 3 years	> 3 years
P3	Man	18-24	United States	1-3 years	1-3 years
P4	Man	35-44	Georgia	> 3 years	> 3 years
P5	Man	25-34	India	> 3 years	> 3 years
P6	Man	35-44	Germany	> 3 years	> 3 years
P7	Man	25-34	Turkey	> 3 years	> 3 years
P8	Man	25-34	Germany	> 3 years	> 3 years
P9	Man	35-44	Israel	1-3 years	> 3 years
P10	Man	25-34	Netherlands	> 3 years	> 3 years
P11	Man	55-64	United States	> 3 years	> 3 years
P12	Man	18-24	Germany	> 3 years	> 3 years
P13	Man	18-24	South Africa	1-3 years	1-3 years
P14	Man	18-24	United States	1-3 years	1-3 years
P15	Man	18-24	Pakistan	> 3 years	> 3 years
P16	Woman	25-34	United States	> 3 years	> 3 years
P17	Man	25-34	United States	> 3 years	> 3 years
P18	Man	25-34	Canada	> 3 years	> 3 years
P19	Man	25-34	India	> 3 years	> 3 years
P20	Man	18-24	China	1-3 years	1-3 years
P21	Man	35-44	Lithuania	> 3 years	> 3 years
P22	Man	35-44	United States	> 3 years	> 3 years
P23	Man	25-34	Austria	> 3 years	> 3 years
P24	Man	other demographic data was missing			

Table 1. Demographic Information of Interview Participants

4 Results

We begin by sharing interview observations and analysis for three key user behaviors exhibited while using SO, namely browsing Q&A, posting questions, and providing answers, followed by the corresponding survey results. Additionally, we present findings derived from the survey on the interrelationships among these three user behaviors as well as participants' attitudes toward the features of SO. In certain instances, we anonymize quotes to protect participant privacy.

4.1 RQ1: Behavior of Stack Overflow Users

In this section, we present the findings of how developers browse, post, and answer ML questions on SO from both interviews and surveys. The goal is to understand the patterns and preferences in engaging with the ML community on SO.

4.1.1 Browsing Behavior in ML Q&A. **Interview**: We first explored how participants seek solutions to ML problems (i.e., browsing ML Q&A), with participants expressing their preferences for different platforms.[9] While 23 participants mentioned their inclination to use SO, one participant indicated a preference for first referring to GitHub to check relevant documentation:

> *"I will examine the GitHub documentation as well as the main code on GitHub since I believe it is easier to comprehend and won't require much time to identify why it is not functioning correctly."* (P18)

We further investigated their search patterns on SO for ML Q&A. The majority of participants (17 out of 24) tend to only rely on external search engines (e.g., Google) and use specific keywords related to their problem rather than using the SO search engine or constructing full sentences.[10] Five out of 24 participants use both search engines (i.e., external search engines and SO search engines). This preference for the external search engine can be attributed to its convenience, as well as its tendency to prominently display results from SO. One common motivation mentioned by our participants for searching ML Q&A is to browse others' answers and gain different perspectives on their problems. For example, P19 stated,

> *"Sometimes, it also occurs that you extract something from one solution, something from another solution, and combining those elements can be helpful."* (P19)

When exploring the usage pattern on SO, our participants generally exhibit friendly behavior and tend to upvote helpful ML Q&A threads. P1 and P6-7 specifically mentioned they would rely on the number of upvotes to find good answers. However, some participants may downvote answers that are unrelated to the questions. Nevertheless, 37.5% participants (P1, P4-5, P11-14, P17, and P23) believe that other users on SO remain strict regarding the formatting, phrasing, and correctness of ML Q&A, for example:

> *"If your question isn't clear, it's going to be very unfriendly to you."* (P14)

Moreover, we explored participants' experiences while seeking solutions to their ML problems on SO and other similar forums.[11] Participants emphasized the usefulness of various platforms, including framework-specific forums (e.g. PyTorch[12]), programming platforms (e.g. GitHub[13], Kaggle[14], etc.), blogs (e.g. towardsdatascience[15], etc), Stack Exchange[16] and professional Slack Channel[17]. Specifically, participants P4, P8, and P13 asserted that these alternative platforms often offer higher-quality ML Q&A experience compared to SO. They emphasized that these platforms cater more specifically to their ML-related issues and offer enhanced accessibility and expediency in obtaining answers. Additionally, participant P16 observed that discussions on other platforms, such as PyTorch, exhibit greater activity, while P7 affirmed that GitHub holds a more prominent and trustworthy position compared to SO.

[9] refer to I1
[10] refer to I2, I3
[11] refer to I4, I5
[12] https://discuss.pytorch.org/
[13] https://github.com/
[14] https://www.kaggle.com/
[15] https://towardsdatascience.com/
[16] https://stackexchange.com/
[17] https://slack.com/

Survey: In the survey, we delved deeper into how participants use SO for browsing behavior within ML Q&A.[18] We found most participants (85.5%) consistently participate in ML Q&A on SO. However, among the remaining respondents, 55.6% of them express their opinions that SO is not a suitable platform for accessing ML-related posts, opting instead for alternative platforms, and 22.2% of them find the ML Q&A on SO to be too specific to address their problems adequately. Moreover, 16.7% of respondents consider the quality of ML Q&A on SO to be insufficient. Interestingly, it demonstrates that ML practitioners on SO may believe that SO is not the best place for ML Q&A and thus choose not to engage in ML Q&A actively, while the majority of users still rely on SO for ML Q&A.

Regarding specific search patterns, only 27.6% of users report previous utilization of the SO search engine, while others mainly use external search engines, such as Google. This finding aligns with the interview observation, wherein 17 out of 24 interviewees only depend on external search engines. Among the subset of users who have used both the SO search engine and Google (23.5%), a majority (65.2%) of them expressed the belief that the SO search engine is not easier to use than Google when seeking ML Q&A on SO. Moreover, when searching with Google, a significant majority of users (84.0%) rely on specific keywords related to their problems, such as framework or library names, or the first line of an error message when conducting their search for answers. Furthermore, users exhibit a preference for upvoting rather than downvoting ML Q&A during their browsing activities, which is also consistent with insights obtained from interviews. The difference between upvoting and downvoting activity has been statistically validated as significant through a paired t-test: the average frequency score of downvoting is 1.57 (on a scale of 1 to 5, with 1 representing never and 5 representing always), while the average frequency of upvoting is found to be 2.92 ($p < 0.001$, $Cohen's\ d = 1.4$).

Additionally, our investigation into user engagement when browsing ML Q&A on SO revealed that users primarily engage with this content for immediate problem-solving purposes, demonstrating a limited inclination to save it for future reference. Only a small proportion of respondents (5.1%) reported using the save function.[19] This finding suggests that users may prioritize resolving their immediate issues over long-term knowledge retention and, consequently, do not perceive a need to save ML Q&A once their immediate problem has been resolved.

> Participants primarily use SO for immediate problem-solving purposes. They seek diverse perspectives for their problems by consulting multiple ML Q&A threads on SO. Some prefer other platforms due to their perceived higher-quality ML Q&A specific to their problems. In general, they exhibit friendly behavior for browsing ML Q&A on SO but find the standards and requirements for ML Q&A are strict.

4.1.2 Posting Behavior in ML Q&A. **Interview**: We then examined participants' experiences and behaviors on SO related to posting ML questions.[20] 23 out of 24 of our participants have posted ML questions on SO and many of them mentioned they feel welcomed when asking ML questions on that platform. Their questions typically revolved around implementation-related matters, such as data manipulation and enhancing model performance. However, they tend to post fewer conceptual questions, such as those about model theory and mathematics, as they believe these questions are better suited for alternative platforms, such as Stack Exchange[21] and Cross Validated[22]. It is

[18] refer to S16, S17, S20-S24
[19] Save: the function located below the downvote button that allows users to track a question/answer for future usage
[20] refer to I6-I11
[21] https://stackexchange.com/
[22] https://stats.stackexchange.com/

noteworthy that many SO users are unaware of this distinction and participants P2-3, P5-6, P8, P10, P13, and P21-22 highlighted this observation.

> *"I discovered another site, Cross Validated, which is also on the same network. And I noticed that a lot of questions about ML were closed on Stack Overflow. And the comment was given that this is not the site to ask these kinds of questions."* (P21)

Regarding the content of the ML questions they post, more than half of the participants emphasized the significance of providing contextual information, data samples, environment information, and reproducible code within their ML questions. However, some participants (P6, P15, P22, and P24) identify challenges in providing relevant datasets, despite their importance for ML questions. As one participant expressed:

> *"Generally, if you don't provide any kind of data, you're gonna get poor responses about your machine learning question."* (P2)

> *"Sometimes it's hard to provide the data of a certain question because you're not allowed to share the data in many cases. While other users need the data to reproduce the problem then it's difficult."* (P6)

Once they submit their ML questions, participants P3, P8, and P12 especially display a tendency to continually edit their questions after submission, aiming to enhance the quality of their questions and increase the likelihood of receiving helpful answers. However, participants P2, P7, P9-10, P12, and P21 expressed their unwillingness to wait for an answer, with one participant stating:

> *"In general, my question is very urgent, and it is causing a delay in my work. If I don't receive a response within the next two or three hours, I will not rely on Stack Overflow for assistance."* (P7)

Survey: To further understand these experiences, we also surveyed participants about their experience with posting ML questions on SO. Specifically, we asked about the types of questions they frequently post, their strategies for formulating these questions, and their experience with posting ML questions.[23] In our survey, 72.8% of the respondents have posted ML questions on SO, 76.5% of them reported they have posted fewer than 10 ML questions on SO, and only one participant who has posted more than 50 ML questions on SO (see Figure 2). The majority of ML questions posted by participants are implementation questions, which aligns with previous research [11]. Additionally, 54.8% of our respondents find conceptual questions challenging to get satisfactory answers for, while 44.0% specify that mathematical questions are also difficult to obtain good answers for on SO. In contrast, only 35.7% of the respondents believe that implementation questions are challenging to get good answers for.

When phrasing ML questions on SO, users deem certain information as essential. Specifically, 92.6% consider a reproducible code snippet as the most necessary element, followed by providing problem context (90.1%) and describing the solutions they had attempted (80.2%). Since ML questions

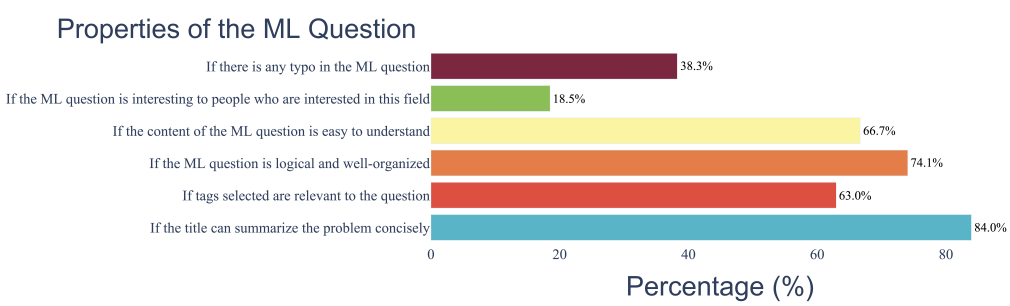


Fig. 1. Properties double-checked by respondents after completing the ML question posts.

[23] refer to S27, S28, S31-S33, S37

often require specific versions of frameworks and libraries to reproduce, the majority of users (69.1%) believe that including environment information is very necessary. Similarly, over half of the respondents (50.6%) value the inclusion of dataset information in ML questions. As mentioned during interviews, the sensitivity of data (which cannot be readily shared in many cases) may affect how much users expect to have a dataset available for ML questions. Table 2 summarizes this. Once they complete their ML question posts, most users (83.9%) make it a point to ensure that the title concisely summarizes the ML problem. They also consider factors such as logical organization, question clarity, and the selection of relevant tags to enhance the quality of their ML questions (see Figure 1). Overall, the majority of participants (80.2%) demonstrate a positive attitude and confidence in formulating ML questions effectively. Most users are open to others modifying their ML questions (72.8%). However, when inquired about their perceptions of the ML community, over half of the respondents reported that they feel excluded (51.9%) and could not receive helpful answers promptly when posting ML questions on SO (53.1%).[24] Specifically, while 63.0% of respondents reported that they could obtain an answer within 2-7 days, only a minimal proportion of questions could be answered within one day (17.3%). When provided with at least one helpful answer, 75.3% of users indicated a preference for the most detailed or comprehensive response, while quicker or simpler answers are less valued by our respondents.

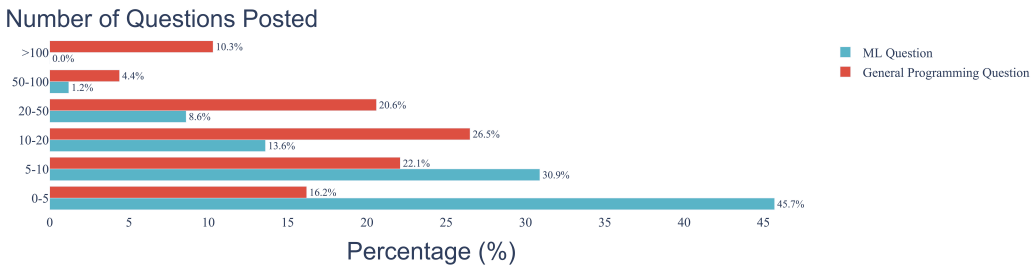


Fig. 2. Comparison of questions posted on Stack Overflow: A breakdown showcasing the distribution of questions posted by both ML practitioners and general users across various categories.

For a more direct comparative analysis, we inquired about participants' posting behavior and perceptions of ML to the broader programming community.[25] Among the respondents who have asked ML questions on SO, 90.0% have also asked general programming questions on SO in the past. The majority of these users possess more experience in posting general programming questions, with 61.8% having asked more than 10 programming questions on SO, and 10.3% having asked more than 100 questions (see Figure 2). A large portion of respondents (89.7%) expressed confidence in formulating well-crafted general programming questions, with 60.7% strongly agreeing. Additionally, 67.6% of respondents do not mind others modifying their general programming questions, which is comparatively lower compared to ML question posters. When comparing the necessary information of general programming questions with ML questions on SO, most respondents agree that problem context, environment information, reproducible code, dataset, and the solutions attempted are more important for ML questions. Furthermore, the properties respondents care about before submitting their general programming questions are similar to those for ML questions, with no significant differences identified. A summary of these findings is presented in Table 2.

In addition to SO, respondents reported other similar platforms to post ML questions, including GitHub issues, Stack Exchange, framework/library-specific forums (e.g., PyTorch forum), and so on

[24]refer to S34-S36
[25]refer to S38-S42

Table 2. Response distributions of the importance of each information in machine learning (ML) questions compared to general programming (GP) questions when posting a question on Stack Overflow. The column *Information* refers to the information respondents may provide in ML questions. The column *Important in ML* refers to the percentage of respondents who consider this information to be important for ML questions. The column *Less important in GP* refers to the distribution of participant agreement for the importance of that particular information in general programming questions compared to machine learning questions, with a Likert scale from disagree to agree, superficial to in-depth, and low to high, respectively. Participants' response to these questions were converted to numeric values from 1 to 5 for analysis and the distribution is shown in the table with (from left to right) a score of one (faint gray), a score of two (light gray), a score of three (gray), a score of four (dark gray), and a score of five (darker gray). The *Average - GP* column lists the average agreement score of the column *Less important in GP. (ML - Machine Learning; GP - General Programming*)

Information	**Important in ML**	**Less important in GP**	**Average - GP**
Problem Context	90.1%	Disagree Agree	3.84
Reproducible snippet of code	86.4%	Disagree Agree	3.69
Solutions original posters have tried	80.2%	Disagree Agree	3.68
Environment information	69.1%	Disagree Agree	3.65
Dataset	50.6%	Disagree Agree	3.71

(see Figure 3).[26] The preference for other platforms stems from their specificity to certain types of questions, such as library-related questions being posted on GitHub issues (75%). This preference is also influenced by the presence of a higher number of ML experts on these platforms (37.5%) and the ability to receive faster answers for their ML problems (32.5%).

> Users may encounter challenges of whether their ML questions can be posted on SO due to suitability issues, which is not clear for the community overall. Relevantly, users have inclusiveness issues when posting ML questions on SO. While comprehensive questions are preferred, it is especially important to include reproducible information for ML questions (though practical constraints exist). In general, users cannot get helpful answers in a timely-urgent manner for ML problems.

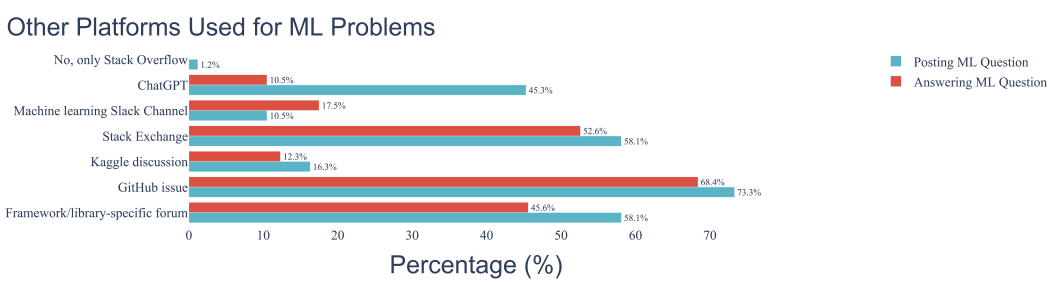


Fig. 3. The percentage distribution of platforms used by respondents for posting and answering ML questions.

[26] refer to S29, S30

Table 3. Response distributions of users' attitudes towards **posting** machine learning (ML) questions and **posting** general programming (GP) questions on Stack Overflow. The column *Agreement* refers to the distribution of participant agreement for that particular statement, with a Likert scale from disagree to agree, superficial to in-depth, and low to high, respectively. Participants' response to these questions were converted to numeric values from 1 to 5 for analysis and the distribution is shown in the table with (from left to right) a score of one (faint gray), a score of two (light gray), a score of three (gray), a score of four (dark gray), and a score of five (darker gray). The *Average* columns represent the average agreement score of each statement. *(ML - Machine Learning; GP - General Programming)*

Post Questions on Stack Overflow	Agreement	Average
ML: I am confident to phrase good ML questions	Disagree Agree	4.05
GP: I am confident to phrase good GP questions	Disagree Agree	4.40
ML: I don't mind others modifying my ML questions	Disagree Agree	3.86
GP: I don't mind others modifying my GP questions	Disagree Agree	3.81
ML: When I am/was a new user, I feel stressed when posting ML questions	Disagree Agree	3.14
GP: When I was/am a new user, I feel less stressed when posting GP questions than ML questions	Disagree Agree	2.59
ML: When I am/was a new user, I feel others are very picky about how ML questions are formatted and described	Disagree Agree	3.26
GP: When I was/am a new user, I feel others are less picky about how GP questions are formatted and described than ML questions	Disagree Agree	2.94
ML: After I become a more experienced user, I feel stressed when posting ML questions	Disagree Agree	2.32
GP: After I become a more experienced user, I feel less stressed when posting GP questions than ML questions	Disagree Agree	3.53
ML: After I become a more experienced user, I feel others are very picky about how ML questions are formatted and described	Disagree Agree	2.35
GP: After I become a more experienced user, I feel others are less picky about how GP question are formatted and described than ML questions	Disagree Agree	3.07
ML: I can get a helpful answer for my ML questions in time	Disagree Agree	3.30
GP: The waiting time for getting helpful answers to my GP questions is shorter than ML questions	Disagree Agree	4.06

4.1.3 Answering Behavior in ML Q&A. **Interview**: Ultimately, we examined participants' behavior in answering ML questions, focusing specifically on their motivations.[27] Among the interview participants, 21 individuals (87.5%) have answered ML questions on SO. Of these, 15 participants (62.5%) mentioned that answering questions enhances their learning process because:

> *"I sometimes answer my questions myself because it's a learning process."* (P23)

[27] refer to I14, I17-I19, I21, I24

Table 4. Response distributions of users' attitudes towards **answering** machine learning (ML) questions and **answering** general programming (GP) questions on Stack Overflow. The column *Agreement* refers to the distribution of participant agreement for that particular statement, with a Likert scale from disagree to agree, superficial to in-depth, and low to high, respectively. Participants' response to these questions were converted to numeric values from 1 to 5 for analysis and the distribution is shown in the table with (from left to right) a score of one (faint gray), a score of two (light gray), a score of three (gray), a score of four (dark gray), and a score of five (darker gray). The *Average* columns represent the average agreement score of each statement. *(ML - Machine Learning; GP - General Programming)*

Answer Questions on Stack Overflow	Agreement	Average
ML: When I am/was a new user, I feel stressed when answering ML questions	Disagree Agree	3.26
GP: When I am/was a new user, I feel less stressed when answering GP questions than ML questions	Disagree Agree	3.14
ML: When I am/was a new user, I feel other users are very picky about how my answer is described and formatted	Disagree Agree	2.51
GP: When I am/was a new user, I feel other users are less critical about how my general programming answer is formatted than ML questions	Disagree Agree	2.39
ML: After I become a more experienced user, I feel stressed when answering ML questions	Disagree Agree	2.68
GP: After I become a more experienced user, I feel less stressed when answering GP questions than ML questions	Disagree Agree	3.25
ML: After I become a more experienced user, I feel other users are very picky about how my answer is described and formatted	Disagree Agree	2.79
GP: After I become a more experienced user, I feel other users are less critical about how my GP answer is formatted and described than ML answers	Disagree Agree	2.93
ML: I feel SO is inclusive when answering ML questions	Disagree Agree	3.51
GP: I feel more inclusive when answering GP questions than ML questions	Disagree Agree	3.30

More than half of the participants expressed a desire to contribute to the community and assist other users. Additionally, participants P20 and P23-24 indicated that answering questions helps them prepare for future interviews. Some participants also emphasized the motivation led by increasing their reputation on SO (Reputation: a rough measurement of how much the community trusts you.[28]). Participants P3, P5, P14, P17, and P22 specifically mentioned that a high reputation can lead to faster responses to their questions. Other participants engage in answering questions purely out of enjoyment or personal interest.

> *"I have a 20,000 reputation now, and users always treat my question with respect. I could imagine if I made a new alias account with a zero score, I would be treated differently."* (P22)

[28]https://stackoverflow.com/help/whats-reputation

Table 5. Motivations and strategies in answering ML questions on Stack Overflow including insights into behavior types, reasons, and preferences. The column *Percent* refers to the percentage of survey respondents who chose this particular reason and respondents are allowed to select multiple reasons at once.

Behavior Type	**Reason**	**Percent**
Why you want to answer ML questions	Give back to the ML community on SO	71.9%
	Help other people	70.2%
	Just for fun	40.4%
	Learn something new	40.4%
How do you look for ML questions to answer	Use tags and SO search engine	59.6%
	Related questions to the questions I got answer on SO	35.1%
	I randomly select ML questions on SO	26.3%
	Rely on the recommendation on SO	21.1%
What ML question do you want to answer	The topics that I am familiar with	75.4%
	The problem that I have experienced before	70.2%
	The question that is easy to understand	63.2%
	The ML question that is easy to reproduce the problem	54.4 %
	The ML question on topics that I am interested in	49.1%
	The ML question with specific background information	40.4%
	The ML question that no one answered	35.1%
	The easy ML question	22.8%
	The challenging ML question	15.8%
	Not duplicate ML question	15.8%

Many of them (P5, P8, P10-11, P13, P15-17, P20, and P23-24) specifically mentioned that having their answers accepted by other users serves as a source of motivation, encouraging them to actively engage in addressing more ML questions on SO.

Having understood their motivations for answering ML questions, we then delved into the criteria by which they select ML questions to respond to.[29] Our participants commonly rely on the tags assigned to each question (Tag: a word or phrase that describes the topic of the question.[30]). They use the SO search engine to locate questions that align with their personal interests or areas of expertise. During the question selection process, the most commonly selected ones are the questions they have successfully resolved in the past, while P3-4, P10, P13-14, P18, P20, and P23 also show a preference to answer questions when they have a high level of confidence in their responses. This tendency may arise from the participants' desire to provide precise and reliable information to other users, which can contribute to the overall quality of answers on SO and uphold their reputation. Moreover, P12, P14, and P16 indicated that they refrain from responding to conceptual questions, as they believe such questions are better suited for Stack Exchange or Cross Validated.

Survey: We further evaluated the answering behavior of respondents in the survey, focusing on their motivation for addressing ML questions on SO compared to other platforms, their preferences for types of ML questions, answering frequency, and the information they prioritize to understand the questions.[31] Among the 74 out of 131 respondents who have answered ML questions on SO,

[29] refer to I15
[30] https://stackoverflow.com/help/tagging
[31] refer to S49-S55

a majority of respondents (73.7%) expressed their intention to contribute to the ML community and help other users by providing answers to their questions. In contrast, some respondents also indicated that their participation is driven by recreational purposes or the desire to acquire new knowledge. Further details can be found in Table 5. In terms of the answering frequency, respondents

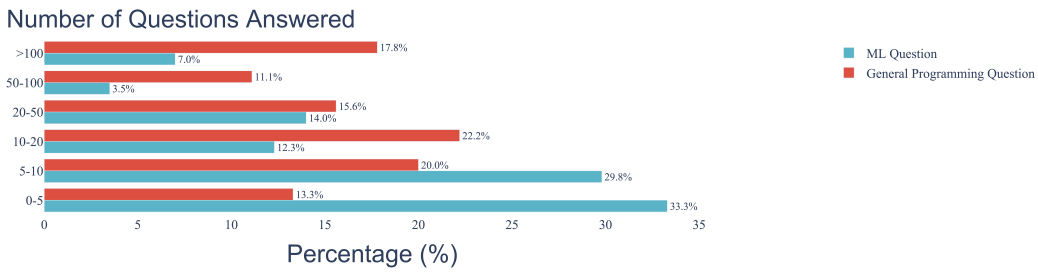


Fig. 4. The distribution of ML and general programming questions answered by respondents.

reported they commonly spend less than an average of 5 hours per week (93.0%) answering ML questions on SO, while the remaining users spend more than 5 but less than 10 hours. Consequently, a significant portion of respondents (63.2%) stated that they have answered fewer than 10 ML questions, whereas 36.8% have responded to more than 10 ML questions on SO (see Figure 4). When seeking ML questions to answer, the most commonly employed approach (63.2%) is to use tags provided by SO through the SO search engine. Another frequently adopted method involves relying on related questions to those they had previously answered or asked on SO. Additionally, 26.3% of respondents reported randomly selecting ML questions they are familiar with to answer, while a smaller proportion of users depend on the recommendations provided by SO. However, half of the respondents exhibit an attitude that SO recommendation is ineffective in enhancing users' awareness about ML questions (see Table 5). Nonetheless, nearly all participants (93.1%) agree that having their answers accepted could serve as a motivator for them to answer more ML questions. When deciding on ML questions to answer, respondents are primarily inclined to choose ML questions related to topics they are familiar with (75.4%), as well as those they have prior experience in. Questions that are easy to comprehend and easy to reproduce are also preferred. A subset of respondents (15.8%) display less concern about answering duplicate questions, despite rules on SO for closing such questions.[32] Table 5 summarizes this. As for the necessary information in ML questions for our respondents to understand the problem, the most crucial aspect, as reported by respondents, is the inclusion of reproducible code or code snippets (85.7%). This is followed by the problem context (78.6%), environment information (71.4%), and the dataset (57.1%). Interestingly, compared to question posters (80.2%, in Section 4.1.2), respondents who answer ML questions tend to care much less (51.8%) about the information of solutions that have been attempted in the description of ML questions, which has been justified as the statistically significant difference ($z = 2.99$, $p = 0.03$, Cohen's $h = 0.50$). Additionally, some respondents (37.5%) expressed that a GitHub repository linked to the problem would be very helpful.

Subsequently, comparisons are drawn between ML and general programming questions when providing answers.[33] Among our respondents who have answered ML questions on SO, 80.7% have also responded to general programming questions on SO previously. Among them, 33.3% have answered fewer than 10 general programming questions, while 66.7% have answered more than 10

[32]https://stackoverflow.com/help/duplicates
[33]refer to S56-S58

(see Figure 4). Moreover, 43.2% of the respondents display a greater inclination to answer general programming questions than ML questions on SO, 34.1% prefer ML questions, and the remaining users have an equal preference for both types. Therefore, it can be inferred that users still possess a relatively strong desire to answer ML questions on SO; however, the response rate and waiting time for ML questions remain relatively low. The challenges faced by users and the possible reasons will be discussed in Section 4.2.3.

Similar to the behaviors of posting ML questions, respondents also indicated the use of alternative platforms to answer ML questions, such as GitHub issues, Stack Exchange, and Kaggle (see Figure 3). The most (77.6%) commonly cited reason for answering ML questions on other platforms is their greater specificity to the ML problems of interest, such as PyTorch-related questions being better addressed on the PyTorch forum. Another reason (20.4%) is that other platforms are more suitable for the long discussion of ML questions, as well as offering a larger number of high-quality questions. Other reasons, including more experts, a friendlier environment, easier questions, and better inclusiveness, are also mentioned but with lower frequency.

> Most respondents spend less than the average of 5 hours per week answering ML questions on SO. When answering ML questions, participants prefer questions they had successfully resolved in the past or had great confidence in their responses. Reproducible code in ML questions is considered more helpful for answering the questions. In addition, question posters with fewer reputations are less likely to be responded to on SO.

4.1.4 Conclusion. In summary, participants often prioritize questions they're confident in answering or have resolved previously when answering ML questions. Reproducible code stands out as a crucial element for effective responses. However, challenges observed regarding responsiveness on SO, particularly for users with lower reputations, lead to uncertainties about question suitability and inclusiveness. While the platform offers a friendly environment, participants emphasized stringent standards for clarity and accuracy in ML discussions. As a result, many may seek alternative platforms, perceiving them to provide more tailored and timely solutions to their specific ML challenges.

4.2 RQ2: Challenges for the Machine Learning community on Stack Overflow

In this section, we discuss the challenges reported by our participants regarding browsing, posting, and answering questions on SO. The goal is to understand the potential reasons for these challenges as well as why the ML community on SO experiences a lower response rate and longer waiting time for questions.

4.2.1 Challenges in Browsing ML Q&A. **Interview**: We interviewed participants about specific challenges or negative experiences they often encounter while browsing ML Q&A on SO.[34] The majority of interview participants indicated that the ML Q&A threads on SO are excessively specific, necessitating adjustments to adapt the provided solutions to their problems. This could be attributed to the absence of essential information (e.g., library version, sample data) or the fast-paced nature of the ML industry (e.g., utilization of outdated or cutting-edge tools). Furthermore, participants P2, P5, P9-10, and P21 highlighted the value of the Comment section, as it offers useful information for clarifying the ML question description or providing improved solutions. When browsing the ML Q&A, they also observe or encounter various common conflicts among users, including arguments, unfriendly comments, questions being closed, and the confusion caused by the conflicting solutions between users, for example:

[34] refer to I4, I5

"All of these rules that Stack Overflow has set up are kind of good rules, but users may take it to an extreme in some sense." (P2)

"They would keep arguing for some time, and I would be there feeling confused and unsure about what solutions to use." (P3)

Despite these challenges, the participants generally perceive the culture of the ML community on SO as friendly, 25% of the participants (P3, P4, P7, P15, P17, P22) specifically concurred that the ML community is more welcoming compared to other communities on SO.

Survey: Building on these perceptions, our survey results further underscore differences in problem-solving within the ML domain compared to others.[35] The results indicate that 64.3% of ML problems can be resolved by reviewing 2-5 Q&A threads, while 31.6% of ML problems require more than 5 Q&A threads. In contrast, individuals who also browse general programming questions on SO indicate that only 21.1% of the questions necessitate more than 5 Q&A threads. This disparity is statistically significant according to the z-test ($z = 2.48$, $p = 0.01$, Cohen$'s$ $h = 0.50$). Additionally, in terms of the quality of ML Q&A, a majority of respondents (75.5%) believe that less than half of the ML answers can be easily reused, which supports the findings from our interviews that respondents often need to adapt the answers to suit their specific problems.

Overall, when browsing ML Q&A threads, only 39.8% of respondents reported never experiencing conflicts, while the most common conflicts encountered are debates between users resulting from differing solutions (31.2%), followed by users reporting to close questions (22.6%), unfriendly comments (16.1%), and communication barriers (16.1%). Moreover, more respondents (62.8%) perceive the ML community's culture on SO to be as friendly as the general programming community, and they feel included when browsing ML Q&A on the platform. However, a larger proportion of respondents (60.0%) expressed agreement rather than disagreement with the notion that the overall culture of the SO community was more friendly in the past than it is now.

> Challenges in browsing primarily include non-reusable ML Q&A, confusion caused by conflicting solutions, and unfriendly comments. Moreover, respondents usually need to browse more threads to resolve their ML problems compared to general programming problems ($p = 0.01$). In addition, 60.0% of respondents agree that the overall culture of the SO had become less friendly over time.

4.2.2 Challenges in Posting ML Questions. **Interview**: When soliciting their perspectives on response times for their ML questions and adverse experiences of posting ML questions on SO,[36] the belief is expressed by 29.2% of participants (P2, P5, P8, P12, P15, P19, P22) that ML questions pose significant challenges in obtaining useful responses when they exhibit excessive specificity (e.g., concerning a particular library/framework version or involving recently introduced tools) or excessive vagueness (e.g., lacking essential information about the environment or reproducible code, subjective questions comparing different technologies, etc.). Furthermore, participants mentioned that if question posters fail to appropriately structure their ML questions or make their question titles/content overly complex, their chances of receiving satisfactory responses are diminished. However, even when individuals adhere to the guidelines provided by SO for posting their questions, half of the participants still perceive the response time to be slow. Additional concerns associated with posting ML questions include unqualified users attempting to provide answers, concerns regarding the tradeoff between the time spent asking questions and the time required to obtain answers, a low rate of accuracy when dealing with complex ML questions, a sense of incompetence

[35] refer to S18, S19, S25, S26

[36] refer to I10-I13

or inadequacy when posting questions, and the perception that other specialized websites offer superior solutions. For instance:

> *"If the problem is too deep in ML, even if you get an answer, the chances of it being correct are usually low."* (P3)

> *"In each step, there might be a problem which is a root cause and you had to describe everything, but you have to do it quickly because if I do it in a long manner describing everything in much detail, then I might end up somebody not reading my questions at all."* (P4)

> *"Whenever I take the time to ask a question, I constantly worry about whether or not people will respond. It's because if they don't, then I feel like I've wasted my time asking the question."* (P9)

> *"They sometimes hesitate to ask questions because it can feel dumb."* (P23)

Survey: To corroborate the insights from the interviews, we also collected the challenges users commonly face when posting ML questions on SO, as compared to general programming questions.[37] Generally, the majority of respondents (69.1%) concur that the duration of waiting time for receiving a helpful response to general programming questions is comparatively shorter than that for ML questions, which has also been revealed in the previous study [11]. According to the survey results, the challenges encountered when posting ML questions on SO can be attributed to several factors. Firstly, ML itself poses difficulties in terms of accurate description, as indicated by 60% of respondents. Secondly, providing the necessary dataset for ML questions on SO proves to be challenging for 56.3% of respondents. Furthermore, 32.5% of participants expressed that the limited time available hinders them from formulating ML questions comprehensively. Additionally, other identified challenges include formatting difficulties (22.5%) and language barriers, such as struggling to understand non-native English speakers or limitations in English proficiency (7.5%).

Furthermore, we explored users' perceptions of the culture of SO and found certain respondents experience stress when they are new to SO and post ML questions. When comparing their experiences in posting programming questions as new users, more respondents (44.1%) agree that they feel stressed as well than disagree (32.4%). As these respondents gain experience, a higher proportion of respondents (56.8%) report a decrease in stress when posting ML questions on SO, while only 17.3% of respondents disagree with this notion. Even though, many experienced respondents (58.8%) still agree that they feel more stressed when posting ML questions compared to general programming questions on SO. Moreover, when our respondents are new to SO, a large proportion (56.8%) also feel that other users are excessively meticulous about the formatting and description of their ML questions, whereas only 24.7% of our respondents hold an opposite opinion. However, as respondents become more experienced, a smaller percentage (39.5%) perceive others as being overly critical of their ML questions, however, there are still more respondents who agree (35.3%) than disagree (22.1%) that other users are less strict in their requirements for general programming questions compared to ML questions. Table 3 summarizes these findings and corresponding statements in the survey.

> The challenges for posting ML questions on SO are primarily in the description of ML questions, limited time, intricate question titles and content, as well as language barriers. Users often perceive the response time to be slower for ML questions. Furthermore, new users often experience stress when posting ML questions, even though this stress tends to decrease as they gain more experience. However, compared to general programming questions, they still tend to feel more stressed when posting ML questions.

[37] refere to S38-S42, S61

4.2.3 Challenges in Answering ML Questions. **Interview**: Lastly, we investigated potential challenges and unfavorable experiences users face when answering ML questions on SO.[38] The biggest challenge for our participants is how to easily find the ML question to answer or raise the exposure of ML problems on SO. Half of the participants (50%) expressed their belief that recommendations on SO are not helpful, particularly in the context of ML questions, which they consider less advisable compared to other general programming questions.

Time constraints also significantly impact our participants. 18 of our participants highlighted that comprehending and responding to ML questions demands a substantial amount of time. Consequently, they sometimes refrain from answering ML-related questions due to the time-intensive nature of these questions. Given the complexity of ML, participants are willing to engage in prolonged discussions with other users to ensure clarity, while it may conflict with their limited available time. Furthermore, 10 participants reported concerns about the quality of questions on SO, such as overly vague, improperly formatted, or expressed in inadequate English, rendering them difficult to understand and, consequently, unanswerable. P21 pointed out that some questions are so rudimentary that users could resolve them through self-search, yet they fail to do so. The presence of an accepted answer, as well as downvotes, significantly diminishes our participants' motivation to respond to ML questions on SO. Specifically, P3, P12, and P14 mentioned that they opt not to answer questions even if the accepted answer is incorrect, for example:

> *"I avoid answering questions marked as accepted because it does not contribute to my reputation. Additionally, if the question is not popular or can be promptly resolved with a correct answer, my response would not be beneficial to me. Hence, it is not worthwhile to invest excessive time in such cases."* (P3)

Moreover, P2, P13, and P22-23 expressed the view that ML questions pose greater challenges compared to general programming questions due to their interdisciplinary nature and incorporation of various mathematical and statistical concepts.

> *"ML questions are harder for me because working with something that is more mathematics than basic logic is a pain."* (P13)

> *"I've been learning machine learning for a while now, but I still feel Like I'm a beginner."* (P22)

Participants also contend that ML questions are inherently challenging due to the difficulty in identifying the root cause of issues (P4, P12, and P20), the challenge of reproducing errors due to variations in library/framework versions (P1, P6, P12, and P16), and the involvement of mathematical concepts (P11, P13). P12 particularly raised concerns regarding the possible theft of answers within the community and P9 proposed the concern that the difficulty of establishing relationships with other users on SO may hinder their motivation to answer questions:

> *"I remember that once I posted some answer, then someone else a few days later posted the exact same answer.* (P12)

> *"Although I frequently encountered certain individuals, I struggled to establish meaningful relationships with them. As a result, I never truly felt part of a cohesive community."* (P9)

Survey: The survey further investigated the challenges respondents face when addressing ML questions compared to general programming questions on SO.[39] The primary challenge (73.2%) faced by respondents is the vague description of the questions. Additionally, question posters often fail to provide sufficient information such as data and environment details needed to reproduce the error (69.6%), and time constraints also limit their ability to answer effectively (69.6%). Other less frequent challenges include difficulties related to mathematics and statistics in ML problems (17.9%) and the complexity of formulating answers for ML questions on SO (7.1%). Despite the challenges,

[38] refer to I16, I19, I20, I22, I23, I28
[39] refer to S49-S58

encouragingly, half of the users (50%) reported never encountering conflicts while answering ML questions on SO. Moreover, the majority of respondents (71.9%) agree that their efforts in answering ML questions on SO can be appreciated by other users. However, some conflicts mentioned by the remaining respondents are similar to those experienced while browsing ML questions, including engaging in debates with other users (23.2%), receiving downvotes (19.6%), and facing unfriendly comments (17.9%). Two respondents specifically noted unresponsive posters who are unwilling to provide further details or fail to respond to answers hinder their motivation to provide more answers.

Except for the external factors, a significant portion of them (57.9%) feel it to be stressful when answering ML questions, especially as new users. This situation improves as they gain experience on the platform, which is similar to the posting activity. However, when compared to general programming questions, a smaller proportion (47.7%) feel stressed when answering general programming questions than ML questions as new users on SO. This attitude persists as they become experienced users. Many of our respondents (45.6%) also feel that others are very picky about how their answers are described and formatted when they are new to SO, which also happens in general programming answers. This situation also changes as they accumulate experience. Fewer respondents (33.3%) indicate other users are picky compared to before (45.6%), while the responses (31.8%) reported others are more critical in ML than general programming questions increase compared to previous (25.0%). Furthermore, respondents display greater confidence in answering general programming questions compared to ML. Half of the respondents (52.3%) feel more assured when answering general programming questions than ML questions. Moreover, a majority of respondents (75.0%) believe that general programming questions require less time to answer compared to ML questions. This perception may stem from the perceived greater difficulty of ML questions, as indicated by most respondents (72.7%). In addition, a majority of respondents (63.6%) agree that the quality of general programming questions is superior to ML questions on SO. In terms of the community culture, half of the respondents (50.9%) reported feeling included when answering ML questions on SO. A higher number of participants (47.7%) express a greater sense of inclusiveness when answering general programming questions on SO compared to ML questions, while only 20.5% disagree with this sentiment. Table 4 summarizes these challenges for answering ML questions on SO. In general, half of respondents (50.9%) do not consider answering ML questions on SO as an important activity for themselves, and a considerable number (33.3%) hold a neutral attitude.

4.2.4 Internal relationship analysis. Based on the findings obtained from quantitative analysis, we specifically investigated the challenges associated with demographic factors and the interrelationships among three specific behaviors (i.e., browsing, posting, and answering ML questions). Our results indicate that Asian and Black respondents are more prone to perceiving significant changes in the culture of SO ($p < 0.001$, $Cramer's\ V = 0.42$), and the proportion of respondents from these two groups who reported a shift in the culture of SO from being more friendly to less friendly is higher compared to respondents from other groups (see Section 6 for more discussion). We also examined factors such as nationality; however, no significant differences were observed. Regarding the respondents' usage behavior on SO, we found that despite users' confidence in formulating high-quality ML questions, they encounter difficulties in obtaining prompt and helpful answers to their ML questions ($p = 0.002$, $Cramer's\ V = 0.41$), which means more confidence is even more likely to lead to a longer waiting time. Conversely, users who post a greater number of questions are more likely to receive answers at a faster pace ($p < 0.001$, $Cramer's\ V = 0.65$). Furthermore, our results indicate that the majority of users consider a waiting period of 2-7 days to be an unreasonable time frame for obtaining answers to their ML questions ($p = 0.005$, $Cramer's\ V$

= 0.35). Furthermore, as discussed in Section 4.1.1, where users often browse ML Q&A on other similar platforms, it is revealed that many users also resort to alternative platforms for seeking or providing ML-related solutions. Besides the commonly cited reason that these alternative platforms are better suited to address specific problems, other major driving factors that lead users towards framework/library-specific forums and Stack Exchange include more ML experts and the expectation of quicker response times. Additionally, users who engage with ML issues on GitHub perceive the platform as more suitable for ongoing discussions related to ML problems (see Figure 5). It is also observed that individuals with more experience in the field perceive ML questions to be more challenging than general programming questions ($p < 0.001$, $Cramer's\ V = 0.41$).

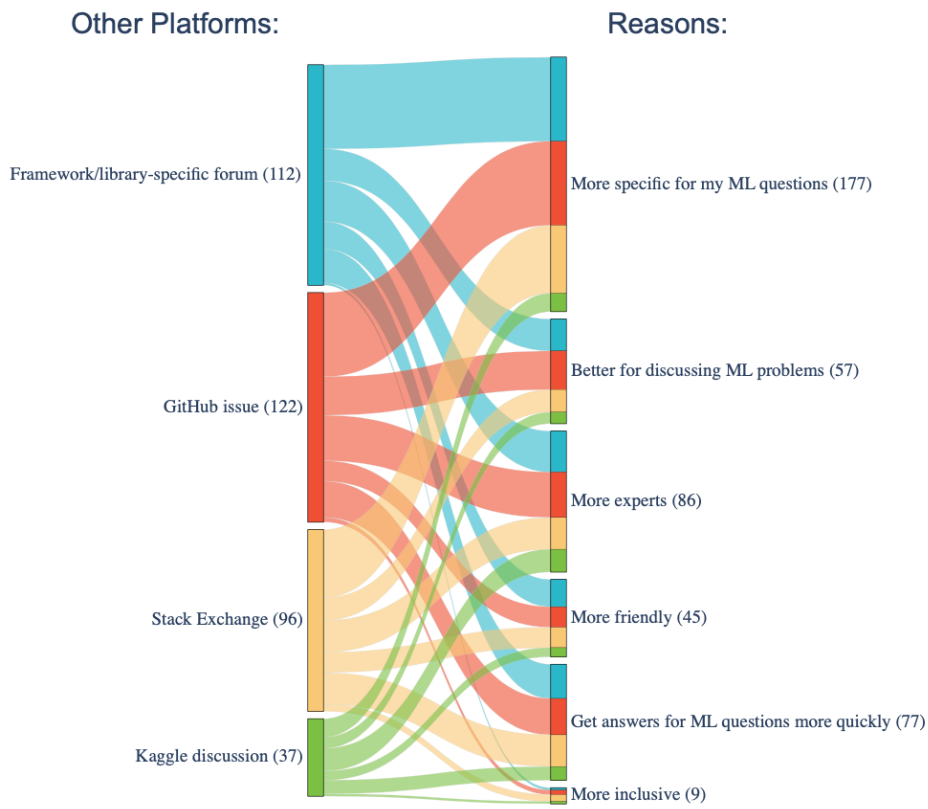


Fig. 5. Other similar platforms commonly used by respondents to post/answer ML questions and the reason why they use them. Numbers in parentheses indicate the frequency of mentions in the survey. Some may be counted multiple times due to the questions allowing for multiple answers.

In addition, we explored the challenges faced by users when attempting to answer ML questions on SO, along with the reasons behind the low engagement in answering ML questions.[40] As shown

[40] refer to S61

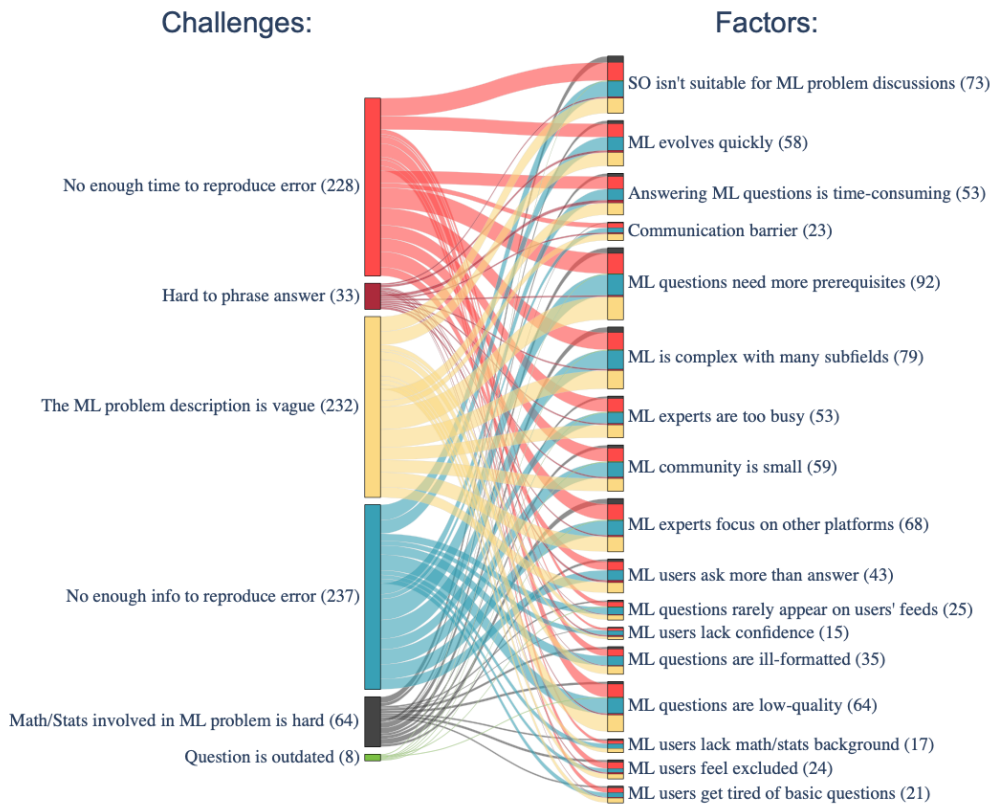


Fig. 6. The challenges encountered when answering ML questions on Stack Overflow and the factors mentioned by respondents that may lead to these challenges. Numbers in parentheses indicate the frequency of mentions in the survey. Some may be counted multiple times due to the questions allowing for multiple answers.

in Figure 6[41], respondents who encounter the challenge of the vague description of the ML questions attribute this limitation more to the absence of prerequisites specific to ML problems, the inherently low quality of ML questions as well as the complexity of ML. For respondents who treat limited time as a challenging factor, they are more likely to consider the factors that SO is not suitable for continuous discussion of ML questions, answering ML is time-consuming, ML experts are too busy as well as more time contributing to other platforms, and also the complexity of ML by nature. The challenge of insufficient information provided ML questions is also mainly because of the complex prerequisites and ML itself as well as the ML questions being hard to format on SO. Respondents

[41]Since the questions in our survey allow multiple answers, the figure illustrates that every node is interconnected. Consequently, we can compare the differences in proportions within each node.

also mention the unique challenges of more mathematics and statistics involved in answering ML questions.

> Challenges in answering ML questions on SO include difficulties in finding ML questions, exposure to ML problems, long time spent answering (general programming questions require less time), question quality, and mathematics/statistics involved in the ML questions. Moreover, users struggle to obtain prompt and helpful answers to their ML questions despite their confidence in asking high-quality questions. Furthermore, newcomers also undergo stress while responding to ML questions, and this stress tends to diminish as they accumulate more experience, although it remains more prevalent compared to general programming questions.

4.2.5 Conclusion. The challenges on SO in the ML community include non-reusable Q&A, conflicting solutions, and perceived declining platform friendliness, with 60.0% noting a less welcoming culture over time. Posting ML questions brings complexities, from crafting precise descriptions to handling linguistic barriers and time constraints. Both newcomers and experienced users report heightened stress with ML questions compared to general programming, compounded by slower response times and difficulties in finding relevant content. Despite efforts to post quality questions, obtaining timely and effective answers to ML challenges remains a persistent issue on SO.

4.3 RQ3: User Suggestions for ML community on Stack Overflow

In this section, we present the desired improvement of the ML community on SO given by our participants from both interviews and surveys.

Interview: The suggestions put forward by the participants during the interview primarily focus on two aspects: enhancing the quality of ML questions and improving the ML community on SO.[42] In particular, participants P8 and P14 both pointed out that the rules on SO are excessively critical and stringent. Participants P10, P15, and P19-20 expressed a desire for the platform to offer more guidance or regulations to help enhance the quality of ML questions, for instance:

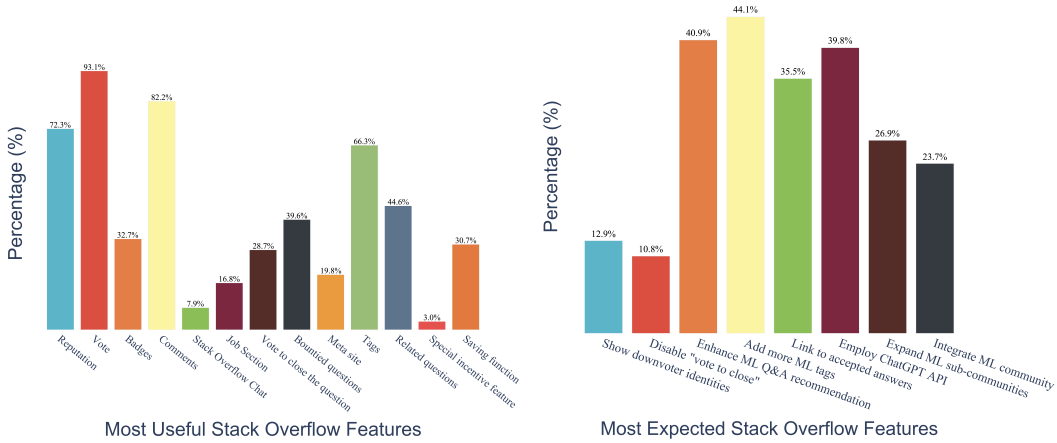
> *"Provide some templates for posting the ML questions, and if you have an issue, there are templates for reference."* (P20)

From a community perspective, numerous participants emphasized the need for a unified community or platform to discuss ML issues. For instance, P7 expressed a desire for SO to be more receptive to conceptual questions. To cultivate a more cohesive ML community, participant P21 emphasized the importance of treating all users equally, irrespective of their experience levels. Additionally, six participants (P5, P7-8, P10, P15, P17) provided suggestions regarding the incentive regulation of SO, hoping for increased incentives to enhance user engagement and attract more individuals to contribute to the ML community:

> *"I feel like there are a lot more people working in the ML domain than people contributing to Stack Overflow. If people could contribute more, the community grows stronger."* (P8)

Survey: In the survey, we also gathered users' feedback and expectations regarding the features of SO.[43] The survey respondents highlighted several noteworthy features of SO based on their usefulness. The most frequently (93.1%) mentioned feature is *vote*[44], followed by *Comments*[45], reputation. On the other hand, the least mentioned useful features are the *special incentive* feature[46], which receive support from only 3.0% respondents, as well as *Stack Overflow Chat* (i.e. chat

[42] refer to I25-I27, I29
[43] refer to S62, S63
[44] https://stackoverflow.com/help/why-vote
[45] https://stackoverflow.com/help/privileges/comment
[46] https://stackoverflow.blog/2022/12/14/hats-out-of-the-bag-join-us-for-winter-summer-bash-2022/



(a) Existing Features of Stack Overflow valued by respondents.

(b) New features of Stack Overflow expected by respondents.

Fig. 7. Perspectives of the respondents regarding the features of Stack Overflow.

rooms)[47], which is supported by 7.9% respondents, and the *job* section, which is mentioned by 16.8% participants (note: job section has been closed permanently). Figure 7a provides more details. Furthermore, participants also expressed their expectations for additional features they would like to see on SO. The feature that received the highest number of votes (44.1%) is the inclusion of more tags to describe ML questions. Better recommendations for ML Q&A interactions rank second, closely followed by the request to import the ChatGPT API, which garnered support from 39.8% respondents. Figure 7b summarizes this. In addition to these findings, participants also offer various practical suggestions to improve SO: (1) When composing a question, provide additional related similar questions displayed in the recommendation center; (2) Integrate the auto-complete feature while writing a question or an answer; (3) Offer user templates that necessitate (particularly inexperienced) users to appropriately structure questions and include an adequate level of information; (4) Downvote wrong comments.

Regarding some general suggestions, respondents raised concerns about certain aspects of SO.[48] They noted that SO follows a "question-then-answer" format, whereas ML questions often require multiple questions and answers for the same original question, attracting multiple parties with different perspectives on the same topic. Respondents suggested that platforms such as PyTorch Forum, which are topic-based or thread-based, may be more suitable in these cases. Furthermore, participants highlighted that in the ML community, many questions may appear to be duplicated but are challenging for beginner users to understand, and the culture of downvoting on sight in SO may not be appropriate for ML-related discussions.

Other suggestions include more incentives to improve user engagement, enhancing ML-specific features and community building, and improving ML question quality and user behavior, for example:

> *"I have never tried to post a question or answer because there is little incentive to answer questions and I assume that I won't receive an answer quickly enough."*

[47] https://stackoverflow.com/help/privileges/chat
[48] refer to S64

"Need more focus on integrating ML-related features such as runtimes for code snippets with data sourcing, and community building."
"The community is too fragmented between Stack Overflow and other Stack Exchange sites. There should be a StackML."
"Be more patient with new people who are learning. Have people explain the problem better and don't assume parts, since the person reading it knows nothing about the particular job of the person posting the question."
"Kindness is key for community building. The easier it is for community members to be kind and the harder it is for us to be unkind, the better."
"A lot of questions are very basic, like someone trying to find someone doing their homework assignment, and often people don't even try to solve a problem but just copy-paste the crash dump and then shout 'solve it!' Sometimes I even gave answers to such questions, but then people did not react. So who knows if this was not useful or not to the person."

Conclusion In summary, respondents emphasize the need for a more active and inclusive ML community, expressing concerns over SO's stringent rules. They advocate for enhanced features such as specific tags and templates on the platform and seek greater receptivity to conceptual ML questions.

Our respondents frequently suggest the creation of a more vibrant ML community, with increased involvement from ML users. They are concerned with the strict rules of SO and wish for more incentives and more responsive users in the ML community. Additionally, they propose enhancing the quality of ML questions with more features, such as more specific tags and templates for reference. They also hope SO to be more receptive to conceptual ML questions.

4.4 RQ4: Impact of ChatGPT

When this study was designed and conducted, generative AI tools (e.g., ChatGPT) were released and raised extensive discussions about their potential impact on programming. Thus, in our study, we also investigated users' perspectives on the impact of ChatGPT on Q&A platforms, specifically for the ML community.

Interview: ChatGPT is a distinctive platform mentioned by our participants for their ML questions, and it does not support interactions among users compared to SO. In discussing ways to improve the ML community on SO, P10 specifically mentioned a shift in his preference due to the emergence of ChatGPT and he prefers the general and comprehensive answers given by ChatGPT when encountering difficulties in ML problems.[49] P2 and P24 also endorsed the general answers for ML questions given by ChatGPT:
"ChatGPT can provide more than just answers [...] you don't have to read the related document and it knows what the function is doing and explains all the variables in it. It saves lots of time." (P10)
"The reason why I might have used ChatGPT is that Stack Overflow gives me solutions for problems, not how to use random forests or how to train." (P24)

P6 and P10 also mentioned their usage of ChatGPT, emphasizing its effectiveness for more swiftly addressing their ML issues compared to SO, which they believe significantly contributes to their productivity.
"If you employ certain techniques and subsequently apply those methods to verify and proceed, it allows you to effectively read the documentation string. This enables you to comprehend the purpose and functionality of the function, along with providing explanations for all the variables involved. Consequently, this approach proves highly efficient, saving a significant amount of time." (P10)

[49] refer to I29

"It's mainly faster than Stack Overflow." (P6)

However, P9 and P22 are also concerned that users are incapable of distinguishing answers generated by ChatGPT from answers written by human developers. Though unknown yet, potential issues may arise if users cannot identify ChatGPT-generated answers.

Survey: Inspired by the interview results, we conducted quantitative assessment of respondents' attitudes toward ChatGPT in the survey.[50] The results demonstrate a high inclination among respondents to seek solutions from the generative AI tools, such as ChatGPT, due to its prompt response time. 42.6% of respondents reported having already used ChatGPT to answer ML questions, among which 58.1% experienced a decrease in their reliance on SO since adopting ChatGPT. Notably, when asked about their perception of ChatGPT, 39.5% of participants believed that it would not replace SO, while 44.2% thought that ChatGPT has the potential to replace SO in the future. Merely 16.3% of respondents remain neutral on this matter.

The utilization of ChatGPT is deemed beneficial by more than half of the users (51.2%) in terms of aiding them in formulating high-quality ML questions on SO while providing detailed descriptions of their issues using proper English. Furthermore, 53.5% of participants believe that ChatGPT could assist them in answering ML questions on SO. Respondents perceive ChatGPT as having a superior ability to address various types of ML questions compared to SO. Specifically, 58.1% of users acknowledged ChatGPT's proficiency in answering conceptual ML questions, while 55.8% indicated its superiority in addressing implementation questions. Furthermore, 58.1% of respondents expressed a clear preference for ChatGPT over SO when confronted with ML problems. This preference may be attributed to the belief held by a significant portion of users (51.2%) that the answers provided by ChatGPT are easier to implement compared to those obtained from SO. However, a majority of participants (55.8%) expressed greater trust in the answers provided by SO in comparison to ChatGPT, while only 16.3% expressed a higher level of trust in the responses generated by ChatGPT. Nevertheless, more than half of the users (58.1%) perceive themselves as being more productive in acquiring ML knowledge when utilizing ChatGPT as opposed to SO.

Conclusion Users believe that ChatGPT can offer more thorough and prompt solutions to their ML issues, but they place greater trust in answers provided by SO. Furthermore, ChatGPT demonstrates the potential to form higher-quality ML questions for SO.

Participants recognize ChatGPT's ability for prompt and detailed ML solutions, though they trust SO answers more. Additionally, ChatGPT shows promise in crafting high-quality ML questions for SO.

5 Discussion

5.1 Response Rate Concerns in Stack Overflow

A notable issue underscored by both our participants and prior research [11] pertains to the low response rate and long waiting time within the ML community on SO, specifically concerning practitioners who require prompt and precise resolutions. Notwithstanding the expansive community and resources available on SO, the delay or lack of responses can hinder users' productivity and growth within the ML domain. To mitigate this challenge, there is a compelling need for a platform that guarantees a faster response rate, thereby enabling users to receive timely support and sustain their progress in their educational and developmental pursuits.

Comparatively, many participants and respondents mentioned the rapid response speed of ChatGPT, though they valued the reliability of answers from SO more. Therefore, it might be advisable to leverage the capabilities of generative language models such as ChatGPT to improve

[50] refer to S59, S60

the quality of both questions and answers within the ML community as well as the response rate. Through its advanced language processing capabilities, ChatGPT can assist users in refining their questions, providing other users with more informative and prompt responses. Additionally, ChatGPT can aid in analyzing and enhancing the general quality of answers timely, especially for conceptual questions, ensuring that the provided information is comprehensive. However, ChatGPT has been banned by SO temporarily due to its low correct rate at the time we drafted this paper.[51] Even though, it is still beneficial to employ external tools to enhance the overall quality of the ML Q&A, as well as to improve response rates and minimize waiting time. We also expect more research on the potential impact and usage of generative AI tools for programming Q&A platforms.

5.2 Improve the quality of ML questions

Based on the aforementioned findings, a significant factor contributing to the diminished response rate is the subpar quality of ML questions on SO. The ambiguous descriptions, absence of reproducible codes, and incorrect use of English in these questions pose challenges for users, hindering their comprehension and ability to formulate appropriate responses. Even though there are many guidelines and rules of SO to help users adapt to the community quickly, it is essential to offer enhanced guidance to ML users, particularly those who are new to ML or SO, in formulating their ML questions better, such as a specific template for ML questions. The enhanced guidance may help articulate their questions clearly and concisely with necessary environment prerequisites, thereby facilitating accurate and relevant responses. Moreover, the quality of ML questions can also be improved by adding more specific features. This includes incorporating more ML-specific tags (e.g. be more specific for the library/framework version, be more specific for the subfields of ML it belongs to, etc.) and establishing connections with external platforms such as GitHub repositories and Google Colab for code sharing, or with alternative platforms such as Pytorch forums, GitHub Issue for more valuable answers for each ML problems. By integrating other platforms, users would be more likely to post ML questions more easily as well as find solutions more quickly, thereby enhancing the overall user experience.

5.3 Build Integrated ML Communitiy

Another prevalent challenge identified in our study is that new users often experience stress when posting and answering ML questions, a phenomenon more pronounced than interactions with general programming questions. This stress could be a contributing factor to the reduced response rates observed within the ML community, as it may pose a constraint to users' engagement. In particular, confusion may occur when new users encounter ML discussions on platforms such as Stack Exchange or Cross Validated, and their ML questions on SO may be reported or closed because other experienced users don't think they belong to SO, which may decrease their motivation or confidence to contribute to ML community or bring stress to their engagement. This observation aligns with the social system theory [63]. ML communities on SO can be viewed as social systems that have evolved their own distinct norms and communication styles. These systems interact with and are influenced by their environment, which encompasses the users, the broader tech community, and the platforms they utilize. New users may encounter challenges due to these established norms and expectations, potentially leading to stress and engagement issues. Furthermore, some users may seek or provide answers on other frameworks or library-specific forums, further decentralizing the ML community. From our study, we also noticed a significant number of ML users who have a background in software engineering or related. Among those individuals who are transitioning their careers or acquiring fresh knowledge, SO emerges as their

[51] https://meta.stackoverflow.com/questions/421831/temporary-policy-generative-ai-e-g-chatgpt-is-banned

preferred platform for learning and sharing programming knowledge including ML. Therefore, it is advisable to achieve an integrated ML community or collaboration among platforms that encourage the participation of ML experts from different domains as well as novices. Additionally, to attract more ML practitioners to this community, it's essential to implement additional incentives that balance the time invested with the rewards earned. Furthermore, fostering a friendly and inclusive environment is crucial, especially given the varying levels of expertise and experience among ML users. This approach not only supports various levels of expertise among ML users but also enhances collaboration, knowledge sharing, and problem-solving across different ML subfields. Such a nurturing environment helps facilitate more open communication and continuous interactions within the ML community. Consistent with Luhmann's theory that communication is fundamental to social systems and catalyzes further interactions [63], promoting ongoing dialogue and feedback within the ML community is essential for its sustained development.

6 Limitation and threats to validity

6.1 Internal Validity

Firstly, the selection of participants and respondents in both studies is based on their activity history under the topic of ML on SO, which might lead to a lack of representativeness across all possible user profiles. For example, brand-new users who have less than one year of experience on SO may not be included in our study. Consequently, the findings of our study may not comprehensively capture the perspectives and experiences of ML users with all possible expertise or experience levels. To mitigate this risk, we included 2 participants without any professional training on programming (i.e. not in any programming-related major) in our interview (P9 and P22). In the survey, we included 6 questions to gather their insights regarding their experiences as new users on SO.

In addition, the extensive number of questions in our quantitative survey could result in respondent fatigue, potentially undermining respondent trust, credibility, and the quality of the data collected. To address these concerns, we organized questions logically and strategically by categorizing questions into relevant sections and presenting them in a coherent sequence, informed by feedback gathered from pilot tests. Furthermore, we integrated interactive features, such as skip logic and branching, to customize the survey experience according to individual respondent characteristics and feedback.

Lastly, we focused solely on SO as the platform for ML Q&A in both qualitative and quantitative studies. However, during our interviews, we found that many users also seek answers to ML questions on other platforms such as Pytorch forum, Stack Exchange, and Cross Validated. Thus, our conclusions and recommendations may primarily apply to SO, and further work might be necessary to generalize these findings to the broader ML community encompassing various platforms.

6.2 External Validity

When determining the ML topics for recruitment in our study, we selected 17 ML-related topic tags (see Section 3). This approach, although adopted from previous research, may not provide a comprehensive and exhaustive sample of all ML questions on SO. Therefore, our findings might be limited in terms of the breadth of ML topics covered.

Additionally, even though we have reached out to over 2,000 users on SO, our study recruited a very small portion of certain demographic groups (e.g., women developers and older developers on SO) for both qualitative and quantitative studies. Thus, our results may not generalize to those specific groups and may not be able to identify unique characteristics of certain demographic groups. Thus, we chose not to present the analysis regarding gender differences in this paper.

However, future research on underrepresented groups (e.g., women) in the ML community on SO is a very important topic for diversity and inclusion in general.

7 Conclusion

As a mainstream programming platform, SO is not primarily designed for ML but has a big ML community for ML learners and experts to share thoughts and seek solutions to problems [44]. The engagement in the ML community on SO experiences a significantly low response rate and longer waiting periods [11], but the underlying reasons remain largely unexplored. In this paper, we presented an empirical study based on 22 hours of semi-structured interviews and 131 survey responses of usage behavior and the challenges of seeking or providing answers for ML questions on SO. We found the main reasons that lead to the low engagement of ML questions include the quality of ML questions, misalignment of time investment and gain, and distributed communities. Compared to general programming problems, ML users have to read more Q&A threads to resolve their problems. Furthermore, despite being experienced users on SO, our participants still experience heightened levels of stress when it comes to posting and answering ML questions compared to general programming questions. We also investigated the influence of ChatGPT for such online question-and-answer platforms, especially for ML Q&A, and proposed the suggestions given by users on SO. We hope that our results can shed light on the ways to help ML knowledge-seeking and sharing and build a more active and inclusive ML community.

8 Acknowledgments

We express our gratitude to the 24 individuals who participated in the interviews, generously sharing their experiences of using SO and providing insightful perspectives. Additionally, we extend our thanks to all the respondents who took the time to complete our surveys.

References

[1] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.

[2] Zeeshan Ahmed, Khalid Mohamed, Saman Zeeshan, and XinQi Dong. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database*, 2020:baaa010, 2020.

[3] Supriya V Mahadevkar, Bharti Khemani, Shruti Patil, Ketan Kotecha, Deepali R Vora, Ajith Abraham, and Lubna Abdelkareim Gabralla. A review on machine learning styles in computer vision—techniques and future directions. *Ieee Access*, 10:107293–107329, 2022.

[4] William R Hersh, Robert E Hoyt, Steven Chamberlin, Jessica S Ancker, Aditi Gupta, and Tara B Borlawsky-Payne. Beyond mathematics, statistics, and programming: data science, machine learning, and artificial intelligence competencies and curricula for clinicians, informaticians, science journalists, and researchers. *Health Systems*, 12(3):255–263, 2023.

[5] Lina Zhou, Shimei Pan, Jianwu Wang, and Athanasios V Vasilakos. Machine learning on big data: Opportunities and challenges. *Neurocomputing*, 237:350–361, 2017.

[6] Nitin Rane. Enhancing mathematical capabilities through chatgpt and similar generative artificial intelligence: Roles and challenges in solving mathematical problems. *Available at SSRN 4603237*, 2023.

[7] Indrajeet Kumar, Jyoti Rawat, Noor Mohd, and Shahnawaz Husain. Opportunities of artificial intelligence and machine learning in the food industry. *Journal of Food Quality*, 2021:1–10, 2021.

[8] KK Ramachandran, A Apsara Saleth Mary, Shibani Hawladar, D Asokk, Bandi Bhaskar, and JR Pitroda. Machine learning and role of artificial intelligence in optimizing work performance

and employee behavior. *Materials Today: Proceedings*, 51:2327–2331, 2022.

[9] Jinfeng Wen, Zhenpeng Chen, Yi Liu, Yiling Lou, Yun Ma, Gang Huang, Xin Jin, and Xuanzhe Liu. An empirical study on challenges of application development in serverless computing. In *Proceedings of the 29th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering*, pages 416–428, 2021.

[10] Alaleh Hamidi, Giuliano Antoniol, Foutse Khomh, Massimiliano Di Penta, and Mohammad Hamidi. Towards understanding developers' machine-learning challenges: A multi-language study on stack overflow. In *2021 IEEE 21st International Working Conference on Source Code Analysis and Manipulation (SCAM)*, pages 58–69. IEEE, 2021.

[11] Moayad Alshangiti, Hitesh Sapkota, Pradeep K Murukannaiah, Xumin Liu, and Qi Yu. Why is developing machine learning applications challenging? a study on stack overflow posts. In *2019 acm/ieee international symposium on empirical software engineering and measurement (esem)*, pages 1–11. IEEE, 2019.

[12] Zishan Qin, Yimin Wu, Jiayan Pei, Jinwei Lu, Shizhao Huang, and Liqun Liu. Related questions retrieval model in stack overflow based on semantic matching. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 321–326, 2022.

[13] Chunyang Chen, Zhenchang Xing, and Yang Liu. By the community & for the community: a deep learning approach to assist collaborative editing in q&a sites. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–21, 2017.

[14] Ke Liu, Guang Yang, Xiang Chen, and Chi Yu. Sotitle: A transformer-based post title generation approach for stack overflow. In *2022 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, pages 577–588. IEEE, 2022.

[15] Md Johirul Islam, Hoan Anh Nguyen, Rangeet Pan, and Hridesh Rajan. What do developers ask about ml libraries? a large-scale study using stack overflow. *arXiv preprint arXiv:1906.11940*, 2019.

[16] Thiago Baesso Procaci, Sean Wolfgand Matsui Siqueira, Bernardo Pereira Nunes, and Terhi Nurmikko-Fuller. Experts and likely to be closed discussions in question and answer communities: An analytical overview. *Computers in Human Behavior*, 92:519–535, 2019.

[17] Nathan Baker, Frank Alexander, Timo Bremer, Aric Hagberg, Yannis Kevrekidis, Habib Najm, Manish Parashar, Abani Patra, James Sethian, Stefan Wild, et al. Workshop report on basic research needs for scientific machine learning: Core technologies for artificial intelligence. Technical report, USDOE Office of Science (SC), Washington, DC (United States), 2019.

[18] Seyed Mehdi Nasehi, Jonathan Sillito, Frank Maurer, and Chris Burns. What makes a good code example?: A study of programming q&a in stackoverflow. In *2012 28th IEEE International Conference on Software Maintenance (ICSM)*, pages 25–34. IEEE, 2012.

[19] Oluwaseun Alexander Dada, George Obaido, Ismaila Temitayo Sanusi, Kehinde Aruleba, and Abdullahi Abubakar Yunusa. Hidden gold for it professionals, educators, and students: Insights from stack overflow survey. *IEEE Transactions on Computational Social Systems*, 10(2):795–806, 2022.

[20] Sunila Gollapudi. *Practical machine learning*. Packt Publishing Ltd, 2016.

[21] Amir Masoud Rahmani, Efat Yousefpoor, Mohammad Sadegh Yousefpoor, Zahid Mehmood, Amir Haider, Mehdi Hosseinzadeh, and Rizwan Ali Naqvi. Machine learning (ml) in medicine: Review, applications, and challenges. *Mathematics*, 9(22), 2021.

[22] Dane Morgan and Ryan Jacobs. Opportunities and challenges for machine learning in materials science. *Annual Review of Materials Research*, 50:71–103, 2020.

[23] Mason McComb, Robert Bies, and Murali Ramanathan. Machine learning in pharmacometrics: Opportunities and challenges. *British Journal of Clinical Pharmacology*, 88(4):1482–1499, 2022.

[24] Anuj Karpatne, Imme Ebert-Uphoff, Sai Ravela, Hassan Ali Babaie, and Vipin Kumar. Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, 31(8):1544–1554, 2018.

[25] Balqis Albreiki, Nazar Zaki, and Hany Alashwal. A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, 11(9):552, 2021.

[26] Andrei Paleyes, Raoul-Gabriel Urma, and Neil D Lawrence. Challenges in deploying machine learning: a survey of case studies. *ACM Computing Surveys*, 55(6):1–29, 2022.

[27] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

[28] Matthias Scheutz. Ethical aspects and challenges for interactive task learning.

[29] Ismaila Temitayo Sanusi, Solomon Sunday Oyelere, Henriikka Vartiainen, Jarkko Suhonen, and Markku Tukiainen. A systematic review of teaching and learning machine learning in k-12 education. *Education and Information Technologies*, pages 1–31, 2022.

[30] Yinong Chen. Iot, cloud, big data and ai in interdisciplinary domains, 2020.

[31] Giang Nguyen, Stefan Dlugolinsky, Martin Bobák, Viet Tran, Álvaro López García, Ignacio Heredia, Peter Malík, and Ladislav Hluchỳ. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*, 52:77–124, 2019.

[32] Natalie Lao. *Reorienting machine learning education towards tinkerers and ML-engaged citizens*. PhD thesis, Massachusetts Institute of Technology Cambridge, MA, USA, 2020.

[33] Joseph Jay Williams, Juho Kim, Anna Rafferty, Samuel Maldonado, Krzysztof Z Gajos, Walter S Lasecki, and Neil Heffernan. Axis: Generating explanations at scale with learnersourcing and machine learning. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*, pages 379–388, 2016.

[34] Kathryn Pole. Mixed method designs: A review of strategies for blending quantitative and qualitative methodologies. *Mid-Western Educational Researcher*, 20(4):35–38, 2007.

[35] Anton Barua, Stephen W Thomas, and Ahmed E Hassan. What are developers talking about? an analysis of topics and trends in stack overflow. *Empirical software engineering*, 19:619–654, 2014.

[36] Sarah Meldrum, Sherlock A. Licorish, and Bastin Tony Roy Savarimuthu. Exploring research interest in stack overflow – a systematic mapping study and quality evaluation, 2020.

[37] Denae Ford, Justin Smith, Philip J. Guo, and Chris Parnin. Paradise unplugged: Identifying barriers for female participation on stack overflow. In *Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, FSE 2016, page 846–857, New York, NY, USA, 2016. Association for Computing Machinery.

[38] Kate Ehrlich, Michael Muller, Tara Matthews, Ido Guy, and Inbal Ronen. What motivates members to contribute to enterprise online communities? In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW Companion '14, page 149–152, New York, NY, USA, 2014. Association for Computing Machinery.

[39] Muhammad Ahasanuzzaman, Muhammad Asaduzzaman, Chanchal K. Roy, and Kevin A. Schneider. Mining duplicate questions in stack overflow. In *Proceedings of the 13th International Conference on Mining Software Repositories*, MSR '16, page 402–412, New York, NY, USA, 2016. Association for Computing Machinery.

[40] Jun Lin. Predicting the best answers for questions on stack overflow. 2018.

[41] Haoxiang Zhang, Shaowei Wang, Tse-Hsun Chen, Ying Zou, and Ahmed E. Hassan. An empirical study of obsolete answers on stack overflow. *IEEE Transactions on Software Engineering*, 47(4):850–862, apr 2021.

[42] Liting Wang, Li Zhang, and Jing Jiang. Duplicate question detection with deep learning in stack overflow. *IEEE Access*, 8:25964–25975, 2020.

[43] László Tóth, Balázs Nagy, Dávid Janthó, László Vidács, and Tibor Gyimóthy. Towards an accurate prediction of the question quality on stack overflow using a deep-learning-based nlp approach. In *ICSOFT*, pages 631–639, 2019.

[44] Iraklis Moutidis and Hywel Williams. Community evolution on stack overflow. *PLOS ONE*, 16:e0253010, 06 2021.

[45] Yla Tausczik, Aniket Kittur, and Robert Kraut. Collaborative problem solving: a study of mathoverflow. pages 355–367, 02 2014.

[46] David Hin. Stackoverflow vs kaggle: A study of developer discussions about data science, 2020.

[47] Pradeep Kumar Roy, Sunil Saumya, Jyoti Prakash Singh, Snehasish Banerjee, and Adnan Gutub. Analysis of community question-answering issues via machine learning and deep learning: State-of-the-art review. *CAAI Transactions on Intelligence Technology*, 8(1):95–117, 2023.

[48] Alaleh Hamidi, Giuliano Antoniol, Foutse Khomh, Massimiliano Di Penta, and Mohammad Hamidi. Towards understanding developers' machine-learning challenges: A multi-language study on stack overflow. In *2021 IEEE 21st International Working Conference on Source Code Analysis and Manipulation (SCAM)*, pages 58–69, 2021.

[49] Siw Elisabeth Hove and Bente Anda. Experiences from conducting semi-structured interviews in empirical software engineering research. In *11th IEEE International Software Metrics Symposium (METRICS'05)*, pages 10–pp. IEEE, 2005.

[50] Ann Blandford, Dominic Furniss, and Stephann Makri. *Qualitative HCI research: Going behind the scenes*. Morgan & Claypool Publishers, 2016.

[51] Stephen Louw. Automated transcription software in qualitative research. In *Proceedings of the International Conference*, 2021.

[52] Virginia Braun and Victoria Clarke. *Thematic analysis*. American Psychological Association, 2012.

[53] Benjamin Saunders, Julius Sim, Tom Kingstone, Shula Baker, Jackie Waterfield, Bernadette Bartlam, Heather Burroughs, and Clare Jinks. Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, 52:1893–1907, 2018.

[54] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. How to do better with gender on surveys: A guide for hci researchers. *Interactions*, 26(4):62–65, jun 2019.

[55] Forrest Shull, Janice Singer, and Dag IK Sjøberg. *Guide to advanced empirical software engineering*. Springer, 2007.

[56] David L Raunig, Lisa M McShane, Gene Pennello, Constantine Gatsonis, Paul L Carson, James T Voyvodic, Richard L Wahl, Brenda F Kurland, Adam J Schwarz, Mithat Gönen, et al. Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment. *Statistical methods in medical research*, 24(1):27–67, 2015.

[57] Judy Robertson and Maurits Kaptein. *Modern statistical methods for HCI*, volume 6. Springer, 2016.

[58] Jan De Leeuw, H Jia, L Yang, X Liu, K Schmidt, and AK Skidmore. Comparing accuracy assessments to infer superiority of image classification methods. *International Journal of Remote Sensing*, 27(1):223–232, 2006.

[59] Donald W Zimmerman. Correcting two-sample" z" and" t" tests for correlation: An alternative to one-sample tests on difference scores. *Psicologica: International Journal of Methodology and*

Experimental Psychology, 33(2):391–418, 2012.
[60] Henry Clay. The paired t-test and the wilcoxon matched-pairs test: Comparing the means of two related groups. *for Nursing and Allied Health*, page 139.
[61] Todd Michael Franke, Timothy Ho, and Christina A Christie. The chi-square test: Often used and more often misinterpreted. *American journal of evaluation*, 33(3):448–458, 2012.
[62] Catherine O Fritz, Peter E Morris, and Jennifer J Richler. Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General*, 141(1):2, 2012.
[63] Niklas Luhmann. *Social systems.* stanford university Press, 1995.

Received July 2023; revised January 2024; accepted March 2024