How did the Emergence of ChatGPT Impact Stack Overflow? – A Literature Review

Daniel Staegemann
Otto-von-Guericke University Magdeburg
Magdeburg, Germany   *daniel.staegemann@ovgu.de*

Mariia Rizun
University of Economics in Katowice
Katowice, Poland   *mariia.rizun@ue.katowice.pl*

Christian Haertel
Otto-von-Guericke University Magdeburg
Magdeburg, Germany   *christian.haertel@ovgu.de*

Matthias Pohl
German Aerospace Center (DLR)
Jena, Germany   *matthias.pohl@dlr.de*

Christian Daase
Otto-von-Guericke University Magdeburg
Magdeburg, Germany   *christian.daase@ovgu.de*

Klaus Turowski
Otto-von-Guericke University Magdeburg
Magdeburg, Germany   *klaus.turowski@ovgu.de*

Abstract

As a consequence of ChatGPT's public release in 2022, software developers and learners of the profession were suddenly provided with a completely new and potentially extremely powerful tool to support them in designing and implementing their applications and answering occurring topic related questions. While, previously, community driven question and answer platforms like Stack Overflow were somewhat unique in their value proposition by providing (the chance for) answers directly geared towards the problems users encountered, they now have an alternative that additionally provides rapid response times. To determine how the emergence of ChatGPT impacted Stack Overflow, a literature review was conducted, exploring its effect on the users' participation behavior and their perception. Further, the corresponding implications are discussed, especially highlighting the responsibility of higher education institutions in this context.

Keywords: ChatGPT, large language model, generative ai, stack overflow, literature review.

1. Introduction

Following the public launch of ChatGPT in 2022, generative artificial intelligence (GenAI) in general and large language models (LLMs) in particular have attracted widespread interest of practitioners and academics alike [11], [36]. Due to their ability to produce sophisticated textual outputs, LLMs are seen by many as a promising new tool to increase productivity and offer entirely new services across numerous domains [8], [18], [39].

However, even though they promise huge potential upsides, their use also comes with

significant challenges. Besides the requirement to craft suitable prompts to increase the likelihood of a high-quality result, GenAI's and LLMs' outputs' correctness is also not guaranteed. Hereby, they are especially prone to so-called hallucinations, where they make up information but present them as factual [21], [34]. Nevertheless, they are highly popular and have a huge userbase. One of the areas where their use has become adopted by many is in programming, be it in educational contexts or for productive use [2], [23], [37], [43]. A source on programming knowledge and for fixing encountered issues that has been very popular for a long time [29] is the question and answer community Stack Overflow [41]. Here, users can ask programming related questions, answer other users' questions, engage in topic related discussions, vote on the quality of other users' contributions, and participate in a like-minded community. While the community-aspect constitutes a significant aspect for some users, many primarily use it as a tool to obtain answers on how to solve their problems and circumvent roadblocks in their software development endeavors without a desire to further engage or contribute themselves [30].

Especially for those, who only aim for quick problem solving, the use of tools such as ChatGPT could be attractive since they promise timely answers to their questions and requests without the need to wait for other users to respond, if those do at all, which is not necessarily guaranteed [29]. Further, because their own lack of knowledge is not exposed to other humans, asking a machine instead of actual humans might also pose a lower barrier to some [6]. Yet, not only experienced contributors are important for platforms like Stack Overflow to strive, but also new users that over time integrate into the community and start shaping them themselves [14]. At the same time, at least currently, the human-generated answers on Stack Overflow have a higher average quality compared to those provided by ChatGPT [24], [47]. Thus, such platforms still have a significant role in the software ecosystem, which is why potential threats to their existence could have significant implications.

Therefore, the goal of this work is to explore how the emergence of ChatGPT impacted Stack Overflow and the way it is used. More specifically, while there are several studies on the topic (cf. Section 2.2), to the best of our knowledge, by now, these have not been brought together to provide a more comprehensive overview of the matter as a whole. To bridge this gap, this will be done by means of a structured literature review (SLR) in this work. In doing so, valuable insights can be gained that can, for instance, help inform policies and design decision for Stack Overflow and similar platforms to highlight and strengthen their value propositions, sensitize educators to the developments, allowing them to respond accordingly, and constitute another facet in the societal discourse on GenAI and LLMs.

To provide guidance and a strong focus to this work and the further performed research activities, the following research question (RQ) shall be answered:

RQ: *How did the emergence of ChatGPT impact the software engineering-oriented question-and-answer-community-platform Stack Overflow?*

To answer the RQ, the remainder of this paper is structured as follows. After this introduction, the conducted literature review is described, and its results are presented. Subsequently, the found papers are analyzed. Then, these findings are discussed and contextualized. Finally, a conclusion is given, and it is outlined how future research could build upon this paper.

2. The Review

For answering the RQ, a SLR was conducted. Hereby, ensuring rigor and reproducibility is particularly important to lend credibility to the results [25], [44]. Hence, in line with common recommendations [7], [31, 32], [45], before the actual search process was started, a detailed protocol was developed to steer the process. In the following, the corresponding steps, the considerations underlying the decisions made, and the respective results are outlined.

2.1. The Review Protocol

To identify the relevant literature, a keyword search was conducted since it allows to find contributions from numerous different outlets based on their thematical fit [44]. For this, numerous scientific databases and search engines were used to facilitate a comprehensive

overview that is not limited by the contents available in just a single source. More specifically, Scopus [16] was chosen as a cornerstone because of its comprehensive coverage across different conferences, journals, and publishers. To complement it, Springer Nature Link (Springer) [40] and Web of Science (WoS) [12] were also queried. Further, due to the focus of the work in the realm of computer science, these three were amended by two domain-specific alternatives, namely, the AIS electronic Library (AISeL) [3] and IEEE Xplore (IEEE) [22]. To search these sources in a purposeful manner with a clear focus, a search term was developed that consisted of two components.

The first part covers this review's focus on ChatGPT. For this purpose, a variety of common relevant terms, names of popular related tools, and different spellings were considered to ensure comprehensiveness.

Search Term - Part 1:
llm OR "large language model*" OR "generative artificial intelligence" OR "generative ai" OR "gen ai" OR genai OR gpt OR chatgpt OR copilot OR "co-pilot" OR mistral OR bard OR claude OR gemini OR perplexity OR bing OR llama

The second part is geared towards the focus on Stack Overflow as the object of investigation in particular and covers two different spelling variants.

Search Term - Part 2:
"stack overflow" OR "stackoverflow"

The two parts were connected through an *AND* operator since both aspects need to be discussed in a conjoint context for a contribution to be relevant to the scope of this paper. However, because the sources used provide slightly different interfaces, it was not possible to always use the exact same configuration for the searches. Instead, the two search terms were applied to the available search fields as depicted in Table 1.

Table 1. The utilization of the search terms

Source	Part 1 used in	Part 2 used in
AIS electronic Library	All Fields	All Fields
IEEE Xplore	All Metadata	All Metadata
Scopus	Article title, Abstract, Keyword	Article title, Abstract, Keyword
Springer Nature Link	Title	Keywords
Web of Science	All Fields	All Fields

Consequently, for instance, the final search term for Scopus was as follows, whereas the corresponding terms for the other sources looked slightly different.

The Complete Search Term for Scopus:
(TITLE-ABS-KEY (llm OR "large language model*" OR "generative artificial intelligence" OR "generative ai" OR "gen ai" OR genai OR gpt OR chatgpt OR copilot OR "co-pilot" OR mistral OR bard OR claude OR gemini OR perplexity OR bing OR llama) AND TITLE-ABS-KEY ("stack overflow" OR "stackoverflow"))

The initial search resulted in 295 records being identified. Hereby, 37 of them came from AISeL, 101 originated from IEEE, 86 were identified through Scopus, 43 were found in Springer, and 28 more were added through the search in WoS. However, due to the use of multiple databases for obtaining the initial set of records, there were several duplicate entries amongst them. After these were removed, 237 unique items remained.

Yet, within this set, there were still numerous entries that did not fit to this study's scope, necessitating further filtering. Thus, following common practices [45], a multi-phase approach for narrowing down the literature to the actually meaningful papers was derived in advance and now executed. This way, the necessary level of efficiency could be ensured without compromising thoroughness and diligence. A visualization of this process is given in Fig. 1.
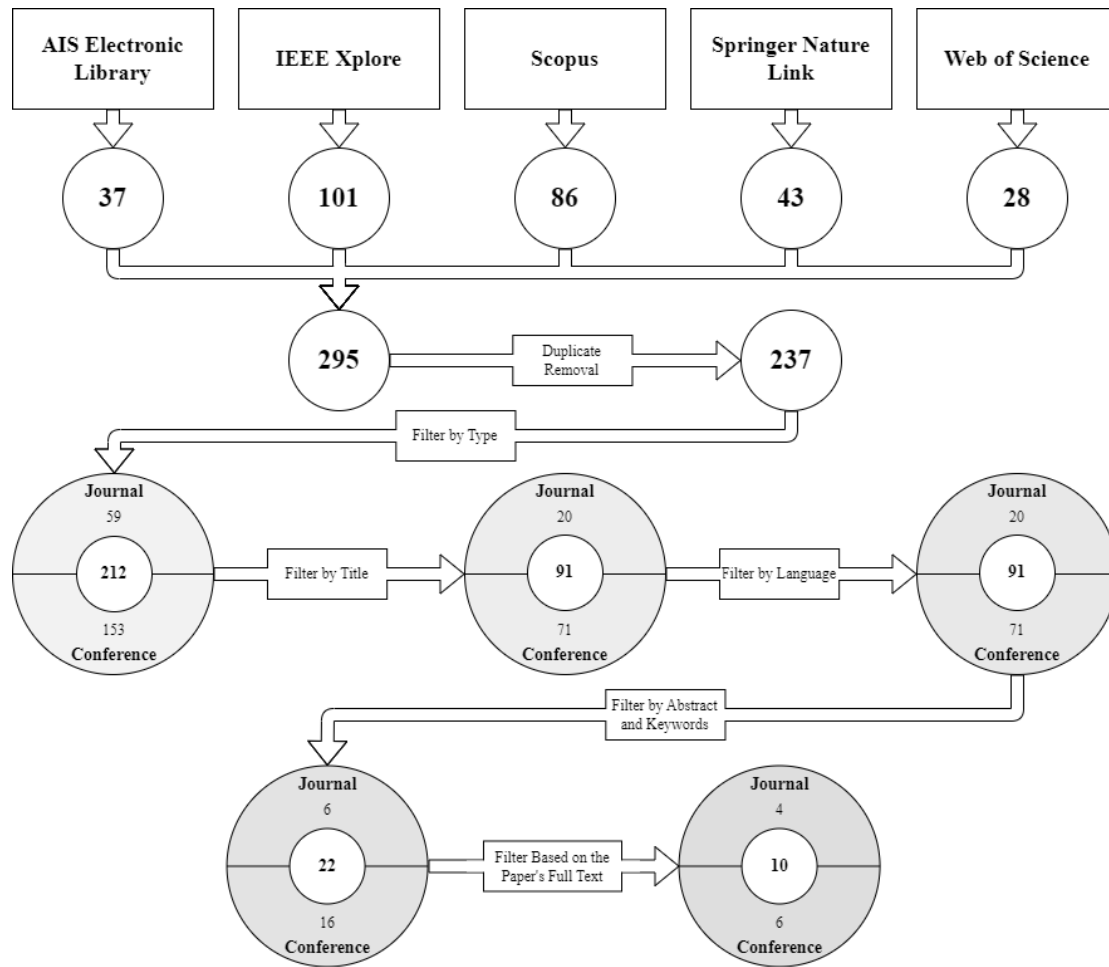
Fig. 1. The search process

To guide the filter process, a set of inclusion and exclusion criteria, as shown in Table 2, was formulated in advance that applied to all of the phases.

Table 2. The search's inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
The paper is written in English	The paper is a duplicate
The paper is published in the proceedings of a scientific conference or in a scientific journal	The found item is a conference review, an editorial, an introduction to a minitrack, a catch word article, a comment, presents the results of a panel or is a similar document type that is no research article in the narrower sense
The paper's focus is on the impact and implications of the emergence of ChatGPT on Stack Overflow	The paper only discusses how LLM-powered tools could be used to enhance the user experience of Stack Overflow
	The paper only compares the quality of human-generated answers on Stack Overflow and ChatGPT-generated answers
	The paper assesses how replacing Stack Overflow as an aid with ChatGPT impacts users
	The paper only discusses the use of data from Stack Overflow to train, test, or benchmark LLMs

Since the RQ aims at the exploration of how the emergence of ChatGPT affected Stack Overflow, this aspect needs to be prominently discussed within a paper for it to be deemed relevant. In contrast, contributions that only focus on aspects such as how LLM-powered tools can be used to enhance Stack Overflow posts or make them more accessible [4], [35], comparisons between the quality of human-generated Stack Overflow answers and ChatGPT-generated answers [27], or the use of Stack Overflow data as input for the training, testing, or benchmarking of LLMs [5], [28] were not considered. The same applies to papers exploring

the implications for users that switch from harnessing Stack Overflow to asking ChatGPT [33]. This is because while they are suitable for exploring the relationship between ChatGPT or other LLMs and Stack Overflow, they do not contribute to answering the RQ.

Additionally, to ensure the necessary rigor and quality, only *conference papers* or *journal articles* were included. Further, within these, only *research articles* were kept, whereas other contributions such as editorials, catch word articles, comments, conference reviews, minitrack introductions, summaries of panel discussions and other similar items were to be excluded. Moreover, due to their common lack of peer-review, book chapters were also not included. For the same reason, preprint-services such as arXiv [13] were also not queried, when the initial search was conducted. While papers that can be found there are sometimes of very high quality and the services' concept is beneficial for timeliness, the lack of peer-review also leads to "concerns about the research accuracy, quality, and credibility of preprints" [1].

Since the metadata that are provided by the searched databases are not always correct and sometimes also lack specificity, the filtering by record type required manual review and, occasionally, adjustment. Therefore, the differentiation between the publication types was only made after this step. Once the documents with unsuitable types were removed, 212 papers remained, of which 153 were conference papers and 59 journal articles.

Following this, the papers were filtered based on their titles. Hereby, when the title made it obvious that a paper was not relevant for answering the RQ, it was excluded, whereas, in cases where it was not absolutely clear if it fits the scope or not, the paper was kept. As a result of this step, 71 conference papers and 20 journal articles remained for a total of 91 records.

Then, to ensure that the authors of the publication at hand could properly understand the contents of the papers, another filter step, based on their language, was prescribed. Hereby, only contributions that were written in English should be kept. However, this applied to all remaining ones, which is why no items were excluded. Nevertheless, this step is still mentioned for the sake of accuracy and completeness, since it was part of the developed review protocol.

Afterwards, the set was further filtered by the papers' abstracts and keywords. Again, when it was not clear if a paper is relevant or not, it was carried over into the next phase. However, this step still led to a significant reduction. Once it was finished, a total of 22 papers remained with 16 being published in conference proceedings and 6 in journals.

Finally, to concludingly assess if these papers are relevant to the scope of this work, they were read in total. Hereby, another twelve items were removed. Thus, for the final literature set, six conference papers and four journal articles remained for a total of ten contributions.

While this is a rather big reduction from the initial number of 295 items that were obtained during the keyword search, this is, besides the occurrence of some duplicates, primarily caused by the specific focus of this work on one direction of the relationship (the impact of ChatGPT on Stack Overflow but not the influence of Stack Overflow on ChatGPT), which can, however not properly be reflected in the search term, as well as the desire to ensure comprehensiveness, even though if it comes at an initially slightly increased effort for the filter process.

2.2. The Identified Papers

An overview of the ten papers that compose the final literature set and, thereby, the foundation for the succeeding analysis is given in Table 3. In addition to an ID, the publication year, the title, the type of paper and the reference, it is also depicted where the respective paper was found. Hereby, it is noteworthy that four of the five used sources, namely AIS, IEEE, Scopus, and Springer, contributed papers that were not found in the other databases. This, in turn, highlights the importance of searching multiple sources to maximize the coverage. Only WoS was not necessary for obtaining the final set, which was, however, not possible to know in advance and its inclusion still increased the search's comprehensiveness.

Further, it is noticeable that all of the papers are from 2023 or 2024. However, this is not surprising considering the time delay that is usually associated with peer-reviewed publications. Since ChatGPT emerged at the end of 2022, there was not much time left for publications, whereas the fact that the review described in the publication at hand was conducted in March 2025, which limits the time that was available for papers to appear in that year.

Table 3. The identified papers

ID	Title	Year	Type	Found in	Ref.
1	Exploring Student Preference between AI-Powered ChatGPT and Human-Curated Stack Overflow in Resolving Programming Problems and Queries	2023	Conference Paper	IEEE; Scopus	[19]
2	ChatGPT Is A User-Generated Knowledge-Sharing Killer	2023	Conference Paper	AISeL; Scopus	[48]
3	Adoption of Artificial Intelligence in Online Communities: A Socio-Technical Perspective	2024	Conference Paper	AISeL	[9]
4	The consequences of generative AI for online knowledge communities	2024	Journal Article	Scopus; Springer; WoS	[10]
5	Large language models reduce public knowledge sharing on online Q&A platforms	2024	Journal Article	Scopus; WoS	[15]
6	"Math is a pain!": Understanding challenges and needs of the Machine Learning community on Stack Overflow	2024	Journal Article	Scopus	[17]
7	Impacts of generative AI on user contributions: evidence from a coding Q &A platform	2024	Journal Article	Springer	[26]
8	Investigating the Relative Impact of Generative AI vs. Humans on Voluntary Knowledge Contributions	2024	Conference Paper	AISeL; Scopus	[38]
9	Investigating Developers' Preferences for Learning and Issue Resolution Resources in the ChatGPT Era	2024	Conference Paper	IEEE	[42]
10	Human Reaction to GenAI Threats: A Perspective of Protection Motivation Theory	2024	Conference Paper	AISeL	[46]

3.   Findings

While the literature on the impact that ChatGPT's emergence had on Stack Overflow is somewhat limited, the identified set still shows that this topic is deemed highly relevant, which is emphasized even more by the fact that already several journal articles exist that are dedicated to the topic. The contributions cover several different aspects, allowing to obtain a comprehensive view, with the corresponding diversity also justifying the creation of the study at hand as a single source, where these different viewpoints and aspects are amalgamated.

While these aspects are not always clearly separated, in general, the findings can be divided into two categories. On the one hand, there is the (quantifiable) impact on user engagement and on the other hand, there are observations regarding the perception and acceptance of ChatGPT in the context of Stack Overflow.

3.1.   The Impact of ChatGPT on the User Activities on Stack Overflow

When ChatGPT emerged, this significantly changed the landscape for (aspiring) software developers because suddenly there was a new tool that has the ability to answer questions, generate code, point out errors, explain concepts, support learning, and much more. Additionally, all of this is (in most cases) provided in real-time, without the need to wait for an extended time until the desired response is available. Hereby, especially this immediate feedback can be highly valuable when an encountered roadblock shall be overcome in a timely manner. Further, this immediacy also allows the user to adjust questions and requests based on the obtained responses should the initially used ones prove ineffective. Therefore, ChatGPT is a very strong competition for Stack Overflow when it comes to the utility as a tool for solving programming and software development related issues. Since, in contrast to other communities as, for instance Reddit, many Stack Overflow users are not primarily interested in its community aspects and, instead, just seek the quickest path to solving their problems [10], it is no surprise that the availability of ChatGPT resulted in a significant reduction of activity on Stack Overflow.

Despite its significance to the community, over the years, the use of Stack Overflow was already declining [15]. However, compared to the time just prior to ChatGPT's release, it is estimated that Stack Overflow's web traffic was pretty much immediately reduced by approximately one million users per day, which corresponds to a reduction of about twelve

percent [10]. Another study [15] reported even higher losses of up to 25 percent. Hereby, these differences are most likely based on the analyzed timeframe, which ended later for the second study with the authors also highlighting that there was a decline over time that stabilized after roughly six months. Compared to the previously occurring "natural" decline, "this decrease represents more than 5 years worth of deceleration in just half a year" [15], emphasizing just how impactful ChatGPT was on the general usage of Stack Overflow.

The decline also shows when looking more specifically at the volume of questions asked to the community. Three of the found studies specifically analyzed ChatGPT's impact in this regard. The one that was published in 2023 determined a reduction of 2.6 percent [48], whereas the ones from 2024 state "decreased demand for answers" [26], respectively a reduction of questions asked of almost 20 percent [10]. This trend of (partially) replacing Stack Overflow with ChatGPT was also found in a survey geared towards the machine learning community, highlighting that it is not limited to pure coding but also to other related domains [17].

Furthermore, the impact also extends to the questions that are asked. Following the emergence of ChatGPT it, several changes to the characteristics of posed questions were observed [48]. Their average length increased by approximately 2.60 percent, whereas their readability was lowered by 2.56 percent and the level of cognition was reduced by 0.62 percent. Therefore, the average question got longer and harder to understand. This, in turn, suggests that for questions that are rather simple, Stack Overflow has been (partially) replaced by ChatGPT as a quicker and more convenient source of information.

Moreover, not only the questions are affected but also the answering behavior of the platform's users. Similar to the questions, the volume of the answers also significantly decreased, which is only partly explained by the reduced number of questions to answer to [26], [46]. However, in contrast to the former, where the average length increased, for the latter, the answers became shorter. Yet, interestingly, the given explanations highly differ. Whereas the authors of [26] suspect the reason to be a certain degree of demotivation caused by the emergence of an increasing number of artificial intelligence (AI) generated answers, in [46] it is assumed that the reason lies in the AI policy of Stack Overflow compared to other somewhat related communities. The hypothesis is that Stack Overflow's ban on the use of GenAI results in the users focusing their efforts on other platforms such as Reddit, where they have to "compete" with the likes of ChatGPT to set themselves apart, while this is not necessary for Stack Overflow.

While the aforementioned considerations are referring to Stack Overflow's userbase as a whole, for a more nuanced view, within [10], [26], and [38], the users were further differentiated based on their activity level [26] respectively account age [10] on the platform. In doing so, the authors of [26] observed that heavy users' contributions showed a decline of 8.1 percent and also showcased a certain drop in quality, whereas they noticed no statistically significant changes for light users. In turn, [10] highlights that following ChatGPT's release, active participants "were increasingly likely to be more established, older accounts" [10]. Mentioned possible explanations are the reduced peer interaction as a possibility to build a network and gain reputation [10], [26], [38] as well as the assumed inexperience of user behind newer accounts to those with a longer history, which might reduce their ability to detect mistakes on ChatGPT's outputs, giving them a heightened (and not always warranted) sense of trust [10]. Emphasizing the former aspect, it was determined that it is "the importance placed on reputation, that primarily drives the demotivation effect" [26].

This heterogeneity also shows when regarding not the users but the topics, where, expectedly, "the most substantially affected topics are those most heavily tied to concrete, self-contained software coding activities" [10], for which ChatGPT is presumably particularly strong in assisting.

Interestingly, since Stack Overflow did not immediately ban the use of ChatGPT, there was a brief period (from its launch on November 30, 2022, until its ban on December 5, 2022), where it could be officially used to generate answers for the platform. The authors of [38] took this as an opportunity to explore the impact the tool made. They collected a dataset of 17916 questions and 60668 answers that originated from 14534 users and used the OpenAI GPT-2 Output Detector to identify ChatGPT-generated responses, assuming sufficient accuracy in the assessment. Of the collected sample, they classified 16715 answers for 3791 questions as AI-

generated. Overall, these questions got 40023 responses. Thus, in total the percentage of AI-generated answers was roughly 27.55 percent, whereas it was about 41.76 percent when only accounting for answers to questions that had at least one AI-generated response. Analyzing these posts and the perception they received in the form of upvotes, the authors determined that "When using GAI tools, answers, on average, receive lower upvote scores, are shorter in length, are easier to read, and are more positive and objective" [38]. Hereby, they explain these findings with ChatGPT's tendency for occasionally producing incorrect answers as well as with the users' motivation for providing high-quality and detailed answers (that are sometimes more critical) to increase their reputation within the platform. Further, in the past, GenAI has been shown to use more objective expressions compared to humans [20], which also showed in this case. Some of these observed effects, namely the higher upvote score for humans and the increased length of the answers, are even stronger for users with a high reputation. At the same time, the differences in positivity and objectivity are reduced, while the readability is not affected [38]. Even though the explored timeframe was rather short, due to the rapidly imposed ban (that was justified with quality issues of ChatGPT answers), the study gave an impression on the implications of unrestricted AI-use in Stack Overflow with regard to the quantity but also the characteristics of the provided answers.

3.2.   The Perception of ChatGPT in the Context of Stack Overflow

Besides the quantifiable effects of ChatGPT's emergence on the volume of questions and answers on Stack Overflow, it is also insightful to consider the perception of (potential) users to fully understand its impact.

Therefore, this perspective was taken in several studies that aimed to explore, through which means (aspiring) software engineers prefer to obtain new insights and solutions for challenges they encountered [17], [19], [42].

A study with students that were asked to solve tasks of varying difficulty levels with the support of either ChatGPT or Stack Overflow revealed that different groups of students also deviated in their preference [19]. Those that had prior experience in application development favored ChatGPT, while Stack Overflow was preferred by web developers. For game developers, in turn, there was no significant difference found. Even though the scope and sample size of the experiment were somewhat limited, and further studies should be conducted to see if the results hold up, it can be seen as an indicator that there is currently no clear consensus on the superiority of either tool and both can provide value based on the situation and preferences (e.g., speed and convenience vs. correctness).

Another survey, this time addressing software developers and computer science students was conducted to explore, which resources they preferred for learning [42]. While the study also covered various resources besides ChatGPT and Stack Overflow, due to the publication at hand's focus, only the findings related to these two options will be discussed here. When it comes to acquiring new software development skills, the participants preferred ChatGPT and similar tools over Stack Overflow, which is probably based on the easy accessibility of its answers. The same sentiment was found when it came to learning about new technologies in software development. For resolving programming issues and errors, the order of preference between these two stayed the same, however, this time the differences were rather small, putting them at almost even levels of perception. Asked how the emergence of AI chatbots will impact question and answer forums such as Stack Overflow, around 60 percent viewed them as a complement, about 34 percent saw them as a replacement and the remaining participants that varying other stances. Those that saw them as a replacement highlighted their speed, immediacy, and the personalized responses. In turn, participants that saw them "only" as a complement acknowledged the positives but also pointed out such aspects as the community in platforms such as Stack Overflow, the superior human creativity for problem solving, and the lacking reliability of the tools as for why they do not believe that the forums will be made obsolete soon.

Specifically geared towards the machine learning community, another study [17] again highlighted the appreciation for ChatGPT's quick response times, which are in a stark contrast to Stack Overflow, where it is not guaranteed to get an answer in a timely manner, if at all. Further, it was appreciated that its answers are specifically geared towards the respective

questions, that its responses are perceived as easier to implement, and that it can help in understanding the contents of relevant documentations in a comfortable manner. However, the participants also displayed more trust in the answers from Stack Overflow, acknowledging ChatGPT's weaknesses and the risks associated with its use. Though, another strength of ChatGPT that was identified is the ability to help with the formulation of suitable questions to solve a problem, thus hinting towards a path, where ChatGPT and Stack Overflow could be jointly integrated into the workflow for overcoming encountered issues. When asked if ChatGPT has the potential to replace Stack Overflow, 44.2 percent agreed, 39.5 percent disagreed and 16.3 percent of the participants were neutral, again highlighting that there is currently great heterogeneity in opinions on the matter.

In contrast to these varying sentiments and perceptions when it comes to the choice between ChatGPT and Stack Overflow to solve problems or acquire knowledge, the stance of Stack Overflow's userbase on the integration of the platform with AI tools and especially an announced partnership with ChatGPT's developer OpenAI seems pretty clear. A study that analyzed user comments on Stack Overflow [9] identified a heavily skeptical climate, with 98 percent of the comments to the announcement of the partnership showing negative sentiment. Therefore, software developers might see ChatGPT as a valuable complement or even replacement, but Stack Overflow members do not wish for their platform to integrate it into the platform itself, voicing concerns such as the susceptibility of AI for errors and the unwillingness to act as an unpaid creator of training data for the AI.

4. Discussion

Even though the number and temporal scope of the studies related to the RQ are still somewhat limited, due to the newness of the topic, many relevant aspects have already been, at least preliminarily, explored, providing a solid initial understanding of the impact of ChatGPT's emergence on Stack Overflow and its target audience.

Overall, the studies show that ChatGPT can be a valuable complement to Stack Overflow that especially shines when it comes to the speed and accessibility of the answers [19], [42]. Yet, regarding their contentual quality, human-generated answers are still perceived as superior [17], [19], [38], [42], which is even more pronounced for experienced and reputed users [38]. Still, as a consequence of the release, the usage of Stack Overflow was significantly reduced, amplifying an already noticeable decline over the previous years [15]. Hereby, especially less demanding requests have been redirected to ChatGPT, benefiting from the quick answers [48]. However, since those questions are usually posed by less experienced developers, this also comes with significant drawbacks. On the one hand, they might lack the capabilities to detect errors and flaws in the obtained answers, making them susceptible to incorporating or learning wrong information [10]. On the other hand, new users are essential for the prosperity of communities such as Stack Overflow [14]. Yet, their influx and involvement are reduced due to the availability of ChatGPT [10]. Moreover, the tool was also observed to have a demotivating effect on more experienced and active users, even though regarding the specific implications, the results were somewhat inconclusive [17], [26], [42]. Reducing their engagement, in turn, threatens the utility of Stack Overflow for those that are seeking answers to complex issues or value the community aspects and the prospect of enhancing their social status [26]. Further, the lack of created high-quality programming related content might also impede the future training of LLMs [15]. At the same time, this use of the created content for the training of models is seen critically by some users anyway, potentially leading to a further reduction in activity if they perceive their efforts to be stolen for the sake of enhancing the GenAI capabilities [9].

Consequently, a potential risk for the software development world lies in the, currently at least to some degree, existing culture of community and mutual support declining, increasing the dependence on technical means for getting help with problems, with their quality also deteriorating over time, due to a reduction in human-generated available training data. Further, new developers' ability and willingness to use official documentation as a source of knowledge, instead of relying on prepared and specific information snippets, directly geared towards their very specific problem might be reduced. This could lead to a situation where they are heavily dependent on LLMs for learning and problem solving, without having the necessary capabilities

to properly appraise the quality of the presented output or to properly consult alternative sources. When it comes to addressing these risks and minimizing the impacts, especially higher education institutions could play an important role by fostering a culture of mutual discussion, support, and collaboration to ingrain these values early, raising awareness of the risks and challenges associated with the harnessing of GenAI in programming but also in general, teaching the students how to properly interact with documentations, explicitly highlighting the value and endorsing the use of platforms such as Stack Overflow, putting emphasis on the understanding of concepts instead of pure results, and designing tasks in a way that cannot be easily cheated by outsourcing them to the likes of ChatGPT, reducing the potential temptation for students to do so at the detriment of enhancing their own abilities.

5. Conclusion

The release of ChatGPT impacted many different domains with one of them being software engineering. Since, suddenly, a new and highly convenient tool for answering programming related questions and obtaining relevant knowledge has been available, the pre-existing alternatives a faced with serious competition. This especially applies to the community-driven question and answer platform Stack Overflow, where users can pose software engineering related questions for others to provide answers and hold corresponding discussions. To understand how the emergence of ChatGPT impacted Stack Overflow, a structured literature review was conducted, exploring its effect on the users' participation behavior and their perception of the tool. Hereby, the former focus covered aspects such as the volume of the created questions and answers but also how their structure and content changed in response to the new environment but also dependent on the respective users' characteristics. Subsequently, the corresponding implications of the found developments and potential risks for the software development ecosystem were discussed, especially highlighting the responsibility and possible avenues of higher education institutions to address and alleviate these looming consequences. However, the study also comes with some limitations that have to be taken into account to properly contextualize the work and its findings. Even though the identified literature covers many different aspects of the matter, and several insights were found across multiple papers, the relatively low numbers of papers and the oftentimes limited scopes of investigation still need to be taken into account when appraising the findings. Moreover, the release of ChatGPT is still relatively recent, which entails that the long-term effects it has on Stack Overflow cannot be known and it could also be possible that some of the observed effects abate or even reverse over time. Therefore, in the future, additional studies should be conducted that cover a longer timespan to validate or maybe also adjust the current observations. Further, the factors leading to a reduction in the number of created answers should be explored more extensively, to get a clearer picture on how platforms could counteract them to stimulate active participation to uphold their value proposition. Additionally, there was only one study found that explicitly considered different specializations of developers, which could, however, allow for a more nuanced understanding if expanded upon. Finally, expanding the scope from only focusing on the impact ChatGPT has on Stack Overflow to examining the relationship in its entirety could yield additional valuable insights to potentially foster a symbiotic relationship between the two.

References

1. Adarkwah, M.A., Islam, A.Y.M.A., Schneider, K., Luckin, R., Thomas, M., Spector, J.M.: Are Preprints a Threat to the Credibility and Quality of Artificial Intelligence Literature in the ChatGPT Era? A Scoping Review and Qualitative Study. International Journal of Human–Computer Interaction, 1–14 (2024)
2. Amos, J.P., Amodu, O.A., Raja Mahmood, R.A., Abdulqudus, A.B., Zakaria, A.F., Iyanda, A.R., Bukar, U.A., Hanapi, Z.M.: A Bibliometric Exposition and Review on Leveraging LLMs for Programming Education. IEEE Access 13, 58364–58393 (2025)
3. Association for Information Systems: AIS eLibrary, https://aisel.aisnet.org/ [27.04.2025] (2025)
4. Bappon, S.D., Mondal, S., Roy, B.: AUTOGENICS: Automated Generation of Context-Aware Inline Comments for Code Snippets on Programming Q&A Sites Using LLM. In: 2024 IEEE International Conference on Source Code Analysis and Manipulation (SCAM), pp. 24–35. IEEE (2024)

5. Beau, N., Crabbé, B.: CodeInsight: A Curated Dataset of Practical Coding Solutions from Stack Overflow. In: Findings of the Association for Computational Linguistics ACL 2024, pp. 5935–5947. Association for Computational Linguistics, Stroudsburg, PA, USA (2024)

6. Boguslawski, S., Deer, R., Dawson, M.G.: Programming education and learner motivation in the age of generative AI: student and educator perspectives. ILS 126, 91–109 (2025)

7. Brendel, A.B., Trang, S.T.N., Marrone, M., Lichtenberg, S., Kolbe, L.: What to do for a Literature Review? -A Synthesis of Literature Review Practices. In: Proceedings of the AMCIS 2020 (2020)

8. Brynjolfsson, E., Li, D., Raymond, L.: Generative AI at Work. National Bureau of Economic Research, Cambridge, MA (2023)

9. Bui, Q.N., Tu, Qiang (John), Liu, H., Aube, J., Chen, H., Song, Y., Hu, J., Lv, Y., Li, L.: Adoption of Artificial Intelligence in Online Communities: A Socio-Technical Perspective. In: Proceedings of the Twenty-Ninth DIGIT Workshop (2024)

10. Burtch, G., Lee, D., Chen, Z.: The consequences of generative AI for online knowledge communities. Scientific reports 14, 10413 (2024)

11. Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., et al.: A Survey on Evaluation of Large Language Models. ACM Trans. Intell. Syst. Technol. 15, 1–45 (2024)

12. Clarivate: Web of Science, https://www.webofscience.com [27.04.2025] (2025)

13. Cornell Tech: arXiv, https://arxiv.org/ [27.04.2025] (2025)

14. Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., Potts, C.: No country for old members. In: Proceedings of the 22nd international conference on World Wide Web, pp. 307–318. ACM, New York, NY, USA (2013)

15. Del Rio-Chanona, R.M., Laurentsyeva, N., Wachs, J.: Large language models reduce public knowledge sharing on online Q&A platforms. PNAS nexus 3, pgae400 (2024)

16. Elsevier: Scopus, https://www.scopus.com [27.04.2025] (2025)

17. Fang, Z., Huang, Y.: "Math is a pain!": Understanding challenges and needs of the Machine Learning community on Stack Overflow. Proc. ACM Hum.-Comput. Interact. 8, 1–35 (2024)

18. Filippucci, F., Gal, P., Jona-Lasinio, C., Leandro, A., Nicoletti, G.: The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges (2024)

19. Garcia, M.B., Revano, T.F., Maaliw, R.R., Lagrazon, P.G.G., Valderama, A.M.C., Happonen, A., Qureshi, B., Yilmaz, R.: Exploring Student Preference between AI-Powered ChatGPT and Human-Curated Stack Overflow in Resolving Programming Problems and Queries. In: 2023 IEEE 15th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM), pp. 1–6. IEEE (2023)

20. Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., Wu, Y.: How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. arXiv preprint arXiv:2301.07597 (2023)

21. Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., et al.: A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Trans. Inf. Syst. (2024)

22. Institute of Electrical and Electronics Engineers: IEEE Xplore, https://ieeexplore.ieee.org [27.04.2025] (2025)

23. Jahić, J., Sami, A.: State of Practice: LLMs in Software Engineering and Software Architecture. In: 2024 IEEE 21st International Conference on Software Architecture Companion (ICSA-C), pp. 311–318. IEEE (2024)

24. Kabir, S., Udo-Imeh, D.N., Kou, B., Zhang, T.: Is Stack Overflow Obsolete? An Empirical Study of the Characteristics of ChatGPT Answers to Stack Overflow Questions. In: Proceedings of the CHI Conference on Human Factors in Computing Systems, pp. 1–17. ACM, New York, NY, USA (2024)

25. Kraus, S., Breier, M., Lim, W.M., Dabić, M., Kumar, S., Kanbach, D., Mukherjee, D., Corvello, V., Piñeiro-Chousa, J., Liguori, E., et al.: Literature reviews as independent studies: guidelines for academic practice. Rev Manag Sci 16, 2577–2595 (2022)

26. Li, X., Kim, K.: Impacts of generative AI on user contributions: evidence from a coding Q &A platform. Mark Lett (2024)

27. Liu, J., Tang, X., Li, L., Chen, P., Liu, Y.: ChatGPT vs. Stack Overflow: An Exploratory Comparison of Programming Assistance Tools. In: 2023 IEEE 23rd International Conference on Software Quality, Reliability, and Security Companion (QRS-C), pp. 364–373. IEEE (2023)

28. Lomshakov, V., Kovalchuk, S., Omelchenko, M., Nikolenko, S., Aliev, A.: Fine-Tuning Large Language Models for Answering Programming Questions with Code Snippets. In: Mikyška, J., Mulatier, C. de, Paszynski, M., Krzhizhanovskaya, V.V., Dongarra, J.J., Sloot, P.M. (eds.) Computational Science – ICCS 2023, vol. 14074, pp. 171–179. Springer Nature Switzerland, Cham (2023)

29. Moutidis, I., Williams, H.T.P.: Community evolution on Stack Overflow. PloS one 16, e0253010 (2021)

30. Mustafa, S., Zhang, W., Naveed, M.M.: What motivates online community contributors to contribute consistently? A case study on Stackoverflow netizens. Current psychology (New Brunswick, N.J.) 42, 10468–10481 (2023)

31. Okoli, C.: A Guide to Conducting a Standalone Systematic Literature Review. CAIS 37, 879–910 (2015)

32. Page, M.J., Moher, D., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., et al.: PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ (Clinical research ed.) 372, n160 (2021)

33. Park, A., Kim, T.: Code suggestions and explanations in programming learning: Use of ChatGPT and performance. The International Journal of Management Education 23, 101119 (2025)

34. Perković, G., Drobnjak, A., Botički, I.: Hallucinations in LLMs: Understanding and Addressing Challenges. In: 2024 47th MIPRO ICT and Electronics Convention (MIPRO), pp. 2084–2088. IEEE (2024)

35. Pudari, R., Zhou, S., Ahmed, I., Dai, Z., Zhou, S.: Aligning Documentation and Q&A Forum through Constrained Decoding with Weak Supervision. In: 2023 IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 346–351. IEEE (2023)

36. Raiaan, M.A.K., Mukta, M.S.H., Fatema, K., Fahad, N.M., Sakib, S., Mim, M.M.J., Ahmad, J., Ali, M.E., Azam, S.: A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. IEEE Access 12, 26839–26874 (2024)

37. Rasnayaka, S., Wang, G., Shariffdeen, R., Iyer, G.N.: An Empirical Study on Usage and Perceptions of LLMs in a Software Engineering Project. In: Proceedings of the 1st International Workshop on Large Language Models for Code, pp. 111–118. ACM, New York (2024)

38. Shan, G., Pienta, D., Thatcher, J.B.: Investigating the Relative Impact of Generative AI vs. Humans on Voluntary Knowledge Contributions. In: Proceedings of the 57th Hawaii International Conference on System Sciences, pp. 7490–7499 (2024)

39. Simons, W., Turrini, A., Vivian, L.: Artificial Intelligence: Economic Impact, Opportunities, Challenges, Implications for Policy (2024)

40. Springer Nature: Springer Nature Link, https://link.springer.com/ [27.04.2025] (2025)

41. Stack Overflow: Stack Overflow, https://stackoverflow.com [26.04.2025] (2025)

42. Tayeb, A., Alahmadi, M., Tajik, E., Haiduc, S.: Investigating Developers' Preferences for Learning and Issue Resolution Resources in the ChatGPT Era. In: 2024 IEEE International Conference on Software Maintenance and Evolution (ICSME), pp. 413–425. IEEE (2024)

43. Tona, C., Juárez-Ramírez, R., Jiménez, S., Durán, M.: Exploring LLM Tools Through the Eyes of Industry Experts and Novice Programmers. In: 2024 12th International Conference in Software Engineering Research and Innovation (CONISOFT), pp. 313–321. IEEE (2024)

44. vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: Proceedings of the ECIS 2009 (2009)

45. vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., Cleven, A.: Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. CAIS 37 (2015)

46. Xiao, X., Chen, K., Yan, Z.: Human Reaction to GenAI Threats: A Perspective of Protection Motivation Theory. In: PACIS 2024 Proceedings (2024)

47. Xu, B., Nguyen, T.-D., Le-Cong, T., Hoang, T., Liu, J., Kim, K., Gong, C., Niu, C., Wang, C., Le, B., et al.: Are We Ready to Embrace Generative AI for Software Q&A? In: 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 1713–1717. IEEE (2023)

48. Xue, J., Wang, L., Zheng, J., Li, Y., Tan, Y.: ChatGPT Is A User-Generated Knowledge-Sharing Killer. In: Proceedings of the Forty-Fourth International Conference on Information Systems (2023)