Abstract – Stack Overflow in the Age of AI: Sentiment, User Engagement and Language Trends

Stack Overflow Analysis Group – npybrozo, ctpascua, cclynes

Goal

Our goal is to analyze how AI tools have influenced Stack Overflow activity. We examined trends in AI-related questions, posting frequency, and user sentiment, over time and across programming language levels.

Stack Overflow is a popular Q&A platform for programmers, widely used to seek help and share knowledge.

Data

We constructed our general dataset by joining four publicly available Kaggle datasets:

Stack Overflow Questions (2008–2022) – Provided full post metadata including title, body, tags, creation date, etc. This was useful to analyze data from the date the page was first launched (2008) up until the release of ChatGPT (2022).

60k Stack Overflow Questions with Quality Rating (2016-2024) – allowed us to analyze data during and shortly after the release of ChatGPT (2022).

Most Popular Programming Languages Since 2004 – Provided relative popularity of several programming languages based on their Google search frequency.

GitHub Programming Languages Data – Provided issue counts across repositories by programming language and year. We didn't end up using this data in our analysis.

The data was generally clean (downloaded from Kaggle) but had some missing values, especially in fields that we didn't use at all, like names, which we replaced with "N/A".

We joined these datasets by aligning on the year and programming language, creating a general database.

We then filtered and sampled our database on topics and time relevance to answer our hypotheses about popularity trends and user sentiment based on language level.

Our four datasets show different kinds of skew that affect our results. The Stack Overflow data is skewed toward earlier years, since site usage has dropped over time (as shown in our analysis). In contrast, the GitHub Issues and Most Popular Programming Languages datasets are skewed toward widely used languages, which have much more data than less common

ones. Many of our analyses also depend on post tags, which can be unreliable since they're often user-generated and can be mislabeled, incomplete, or inconsistent.

Model and Evaluation Setup

We used three different analyses to test our hypotheses. For sentiment analysis, we applied the HuggingFace DistilBERT sentiment classification model to Stack Overflow post text. Posts were grouped by programming language level (low vs. high), and we ran an independent two-sample t-test to compare their average sentiment scores.

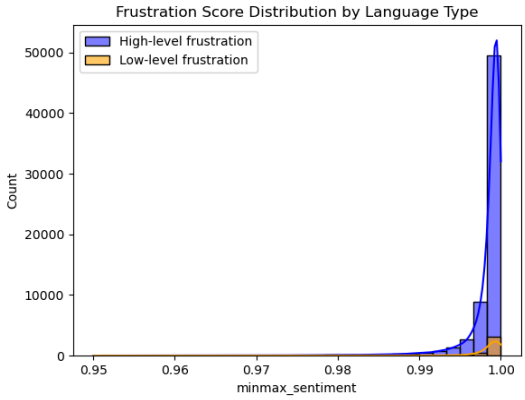
To track AI trends, we identified AI-related questions using keyword matching (e.g., "chatgpt", "llm") in the post body, title, and tags. We then compared the proportion of AI-related posts before and after ChatGPT's release using a chi-square test for independence.

Finally, to analyze usage trends, we aggregated monthly post counts and examined changes in overall question volume before and after ChatGPT. We used linear regression to compare trends over time and a t-test to evaluate differences in average monthly posting rates between the two periods.

Results and Analysis

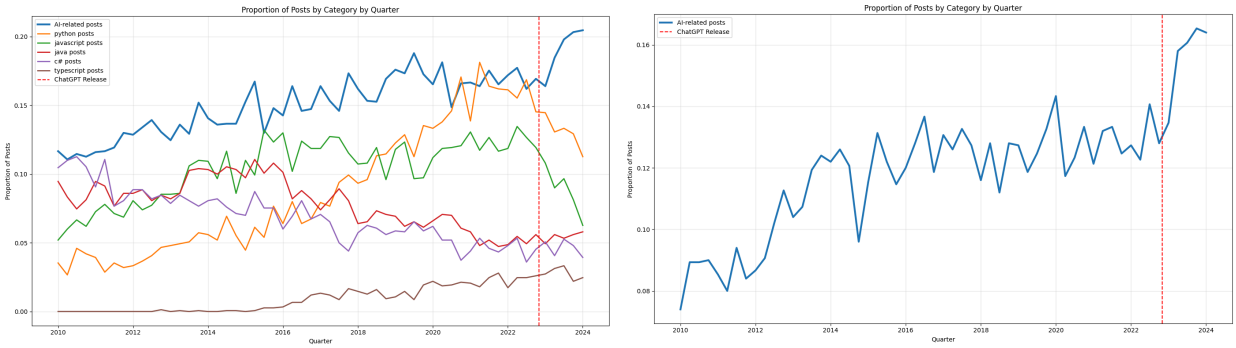
Claim 1: There is no significant difference in sentiment between low- and high-level language questions on Stack Overflow.

Support for Claim 1: The **independent two-sample t-test** produced a p-value of 0.55, indicating we cannot make a conclusion about differences in frustration scores. Limitations included the sentiment model's poor fit for technical text, having been trained on movie reviews.



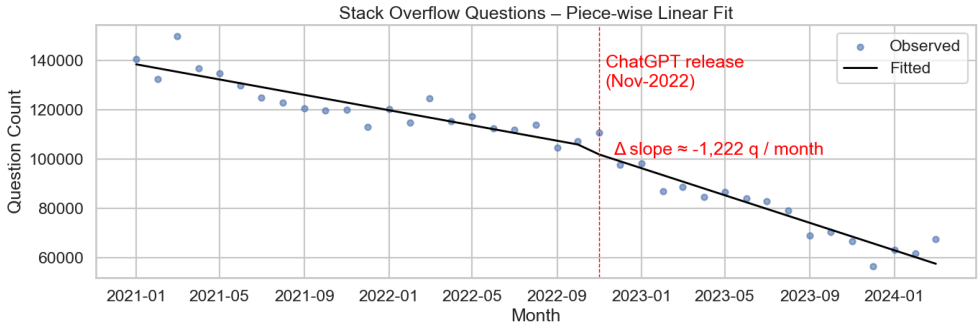
Claim 2: AI-related questions on Stack Overflow increased significantly after ChatGPT's release.

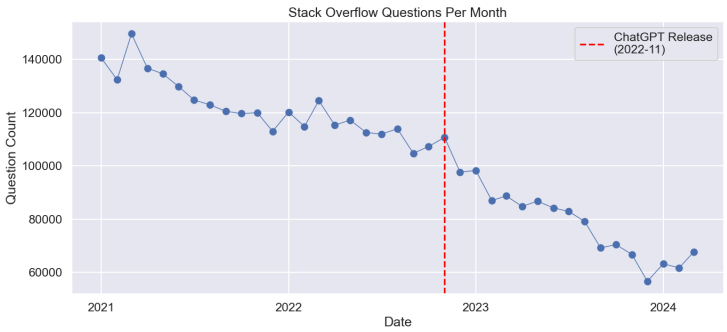
Support for Claim 2: The **chi-squared test** showed an 11.6% relative increase in AI-related content ($p = 0.00076996 < 0.001$) between the pre- and post-ChatGPT periods, based on keyword-matched classification.

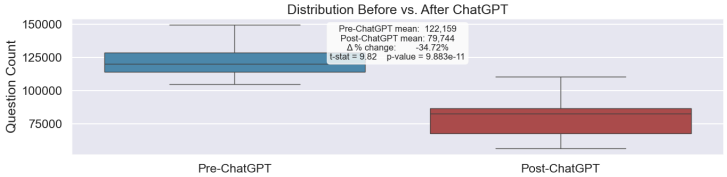


Claim 3: Stack Overflow use declined more sharply after ChatGPT was released.

Support for Claim 3: The regression and **t-tests** showed a steeper decrease in question volume after Nov 2022, the date of release of ChatGPT ($p = 9.883 \cdot 10^{-11} < 0.001$). The decline accelerated compared to the pre-ChatGPT period, supporting the idea that users are turning to AI tools over traditional Q&A forums.







Socio-historical Context

Since the release of ChatGPT, the way people acquire information has shifted dramatically. Many now turn to large language models as their primary source for answers to technical questions. Recent studies note the major impact that this change has had on traditional Q&A forums like StackOverflow and Reddit. For instance, one study from last year reported a “25% decline in StackOverflow activity within just six months of ChatGPT’s launch” and “an overall 50% drop in traffic, questions, and answers” over the past 2 years (Kabir et al., 2024). That finding aligns with our analysis of Hypothesis 3, in which we observed a comparable downward trend in the number of StackOverflow questions — a trend that significantly accelerated after the release of ChatGPT.

Other research suggests that the shift described in our results might have problematic, or at least transformative, consequences as public knowledge-sharing spaces like StackOverflow are “transferred from public repositories to privately-owned AI systems” (Del Rio-Chanona, Laurentsyeva, & Wachs, 2024). One article even describes this as a “cultural change: the collaborative spirit of sharing knowledge is being replaced by the convenience of quick, private solutions offered by AI tools” (After Two Years of ChatGPT, Stack Overflow is Falling Apart – and That’s Terrible News, Code & Hack, 2024). The trend also presents a challenge for the future of large language models, since LLMs also feed on reliable human-generated sources of information to improve their responses. Researchers call this the “reuse paradox” (Code & Hack, 2024). For now, LLMs responses are still highly flawed: “52% of ChatGPT answers contain incorrect information and 77% are verbose. Nonetheless, our user study participants still preferred ChatGPT answers 35% of the time due to their comprehensiveness and well-articulated language style” (Kabir et al., 2024). Thus, while time will tell what the long-term results will be, it is possible that language models and public forums such as StackOverflow will fall into a sort of equilibrium, given the LLMs’ ultimate reliance on StackOverflow to provide a high quality of service.

Ethical Considerations

One potential area of underlying bias derives from the fact that StackOverflow requires answers to be posted in English, even though people use popular programming languages throughout the world. This means that StackOverflow posts likely contain a far higher proportion of English speakers than the true proportion of English-speaking programmers. This might result in an interpretative bias, as well: We are all English speakers residing (at least currently) in the United States, which could lead us to over-generalize the significant trends we found here to other parts of the world, when in reality they might only be present to the same degree among English-speaking Americans. For instance, even though we notice a significant increase in the proportion of AI-related posts, this does not necessarily imply that a greater fraction of people around the world are asking about AI; only that a greater fraction of English communicators are asking about AI on StackOverflow. To prevent this misinterpretation of our results, it is important to acknowledge the limitations of using this StackOverflow data, as we are doing here.

Despite these potential shortcomings, we obtained, aggregated and analyzed the data in a manner that the individuals who provided it would likely agree with. That is because StackOverflow is an explicitly public form, so anyone who asks a question is, by nature of asking, intending for the public to see the question. Furthermore, outside of the sentiment calculations (which are done under the hood), our analysis does not identify specific posts, further anonymizing data that is not even required to be anonymized. The main potential privacy consideration here is the collection of post-specific data by the creators of the sentiment model; however, since the functions used for the sentiment analysis are called and run locally, the third-party creator does not receive this information. The main ethical concerns associated with our project are thus bias- rather than privacy-related.

Citations

On Stack overflow use decreasing after the release of ChatGPT:

Is Stack Overflow Obsolete? An Empirical Study of the Characteristics of ChatGPT Answers to Stack Overflow Questions

After Two Years of ChatGPT, Stack Overflow is Falling Apart — and That's Terrible News

The Decline of Stack Overflow

On public knowledge sharing:

Are Large Language Models a Threat to Digital Public Goods? Evidence from Activity on Stack Overflow

New Study Reveals Impact of ChatGPT on Public Knowledge Sharing

The Decline of Stack Overflow: How AI is Changing Developer Help-Seeking Behavior

Stack Overflow Will Charge AI Giants for Training Data

On language policies:

Stack Overflow Language Policy

Quotes used:

"25% decline in Stack Overflow activity within just six months of ChatGPT's launch [...] Over the past two years, the platform has experienced an overall 50% drop in traffic, questions, and answers." Source

"These numbers highlight more than just a shift in how developers find answers. They signal a cultural change: the collaborative spirit of sharing knowledge is being replaced by the convenience of quick, private solutions offered by AI tools." Source

"This shift introduces what researchers call the "reuse paradox." Models like ChatGPT rely on human-generated content from platforms like Stack Overflow to improve their accuracy. However, as more users rely on AI tools and fewer contribute new knowledge to forums, the flow of fresh content decreases. This lack of new input limits the models' ability to improve and increases the risk of errors or hallucinations in their responses." source

"The site has been a goldmine of knowledge" source

“Rather than completely replacing Stack Overflow, AI is likely to complement it. Developers might rely on AI for quick fixes and basic coding assistance while still using Stack Overflow for more complex discussions and unique challenges. In the future, we might even see AI-powered enhancements within Stack Overflow itself, where AI helps improve answers or guide users in forming better questions.” Source

“52% of ChatGPT answers contain incorrect information and 77% are verbose. Nonetheless, our user study participants still preferred ChatGPT answers 35% of the time due to their comprehensiveness and well-articulated language style. However, they also overlooked the misinformation in the ChatGPT answers 39% of the time.” Source

“The widespread use of AI tools raises concerns about over-reliance on automatically generated answers, which are often incomplete or inaccurate. This dependence can negatively impact the quality of code produced and hinder the development of critical problem-solving skills, especially for new programmers.” source

“[...] since users interact privately with the model, these models may drastically reduce the amount of publicly available human-generated data and knowledge resources.” Source

“These results suggest that more users are adopting large language models to answer questions and they are better substitutes for Stack Overflow for languages for which they have more training data.” Source

“Users may become less inclined to contribute to open knowledge platforms as they interact more with LLMs like ChatGPT, resulting in valuable data being transferred from public repositories to privately-owned AI systems” Source

“[...] questions related to fundamental programming concepts (such as lists, dictionaries, loops, strings, and functions) and data analysis (including pandas, dataframes, arrays, SQL, and NumPy) have experienced the most significant declines. [...] topics related to operating systems and certain development frameworks, IDEs and cloud platforms have experienced comparatively smaller decreases in activity” source