

The Impact of LLM-Based Tools on User Engagement and the Quality of Q&A Discussions on Tech Forums

Bridget Agyare, Sunwoo Baek, Himnish Jain, Eisha Peyyeti

ILLINOIS

Motivation:

StackOverflow is a Q&A forum, used mostly by software developers around the world for problem-solving and knowledge sharing. With the increasing availability of large language models (LLMs), it is crucial to understand their impact on technical Q&A platforms. By tracking interactions from ChatGPT's late-2022 debut through 2024, we expose these shifts and underscore the need to safeguard authentic knowledge exchange.

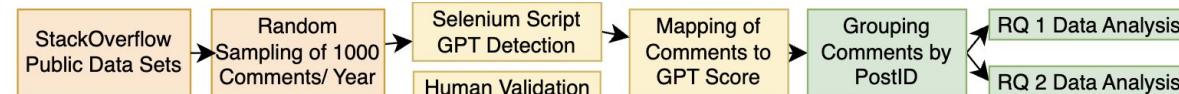
Research Questions:

RQ1: Have users become less engaged, relying more on AI tools instead of participating in discussions?

RQ2: Are answers increasingly AI-generated, and does this impact their quality and perceived helpfulness?

Methodology:

We sampled 1,000 random comments from each year (2022–2024) in the StackOverflow public data dump. Using a GPT-score classifier, threshold-based filtering, and sentiment analysis, we quantified how AI-generated content relates to perceived helpfulness. Finally, we fitted a multivariate regression incorporating four key predictors to predict comment upvote counts.



Findings:

Multiple linear regression:

$$\text{Score} = \beta_0 + \beta_1 \cdot \text{PostLLM} + \beta_2 \cdot \text{Sentiment} + \beta_3 \cdot \text{LogCommentOrder} + \beta_4 \cdot \text{IsHumanWritten} + \epsilon$$

Dependent Variable: comment upvotes (Score)

Independent Variables: pre- or post- LLM (PostLLM), sentiment (Sentiment), comment timing (LogCommentOrder*), and whether the comment was human-written (IsHumanWritten)

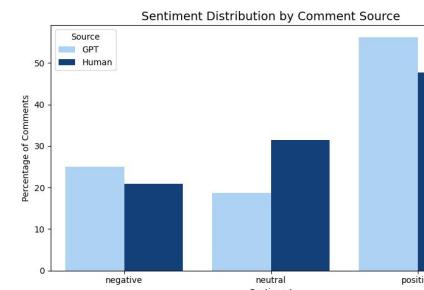
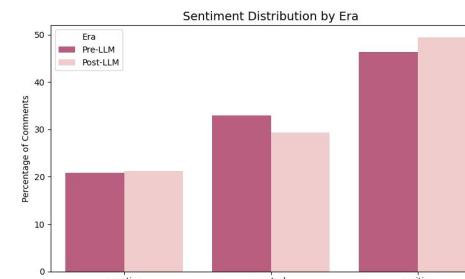
Variable	Coefficient	Std. Error	p-value
Intercept	0.347**	0.106	0.001
PostLLM	-0.066*	0.032	0.042
Sentiment	0.002	0.041	0.953
LogCommentOrder	-0.022	0.291	0.939
IsHumanWritten	-0.049	0.106	0.642

Note: $N = 2000$, $R^2 = 0.002$, Adj. $R^2 = 0.000$.

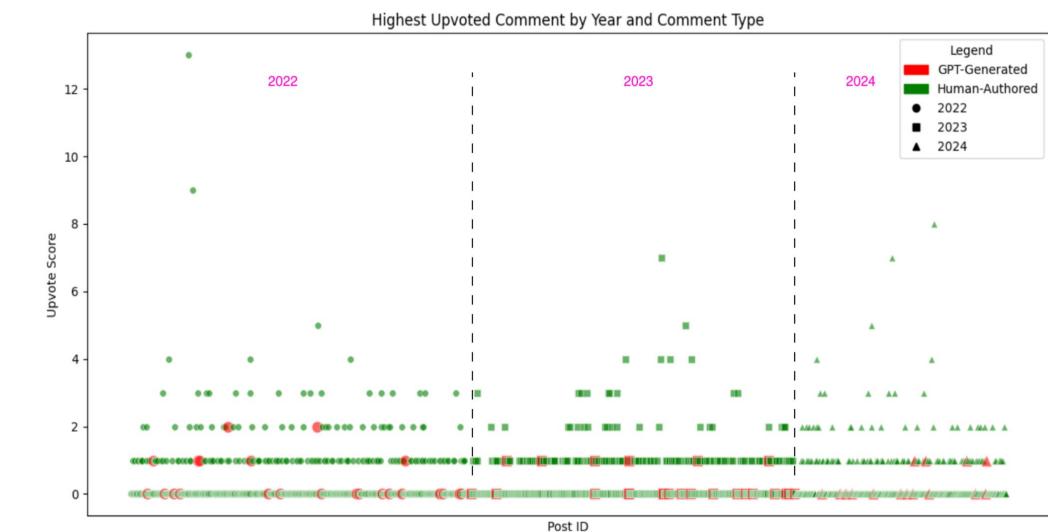
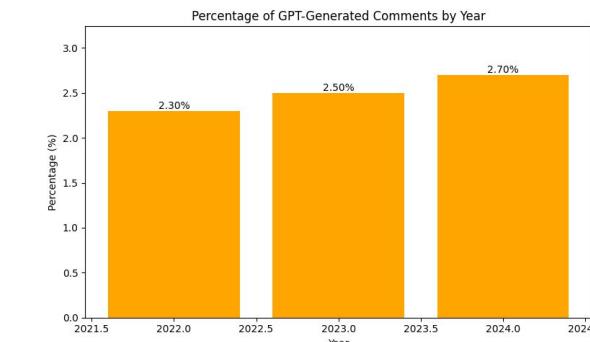
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

- Post-LLM comments received fewer upvotes ($\beta = -0.066$, $p = 0.042$)
- Sentiment, timing, and human-written status were not significant predictors

Sentiment Analysis:



- Sentiment distributions remained stable across eras and comment sources, i.e. there were no significant differences



- **Top upvoted answers tend to be human-written**, indicating community trust
 - Even when GPT-generated comments receive the most upvotes on a post, they **rarely exceed 2 upvotes**, compared to a 4-10 upvote for top human comments

Implications: While AI-generated content on Stack Overflow remains a minority, its presence is growing. We found no statistically significant differences in upvotes or sentiment between human- and AI-generated comments, though descriptively, human-written responses tended to receive more upvotes. As LLM usage expands, it is important to monitor how these dynamics evolve to ensure quality and engagement on technical Q&A platforms.

- LLM generated content on StackOverflow is **steadily rising since 2022**, but represents a minority of the total comments on the platform
 - Detectors like GPT Zero may struggle to keep pace
 - False negatives