The Generative AI Divide: A Descriptive Analysis of Heterogeneous Adaptation Among Knowledge Contributors

ANONYMOUS AUTHOR(S)

How are engagement patterns within online knowledge communities changing in the context of generative AI? While prior research has documented an overall decline in platform activity following the release of ChatGPT, less is known about how different types of contributors have responded to this disruption. In this study, we examine shifts in contributor engagement within online knowledge communities, using Stack Overflow as our focal case. We cluster 394,295 users into canonical contributor profiles based on pre-ChatGPT behavioral data. Using a two-step method combining graph-based community detection and guided Latent Dirichlet Allocation, we identify nine user roles and track their post-ChatGPT engagement. A difference-in-differences analysis comparing each cluster with a matched reference group suggests bounded heterogeneity within a context of platform-wide decline. All contributor types exhibited reduced engagement, though the magnitude varied. Even the most resilient users, such as Steadfast Contributors, showed measurable declines, albeit comparatively milder ones. Other groups, including Peer-Conscious users, Solution Seekers, and Thread Revivers, experienced significantly steeper declines in both questions and answers. These findings suggest that while heterogeneity exists in how contributors responded, it is constrained: no group escaped the downward trajectory. We discuss implications for the evolving dynamics of knowledge sharing as generative AI reshapes participation across online communities.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**.

Additional Key Words and Phrases: generative AI, online communities, knowledge sharing

ACM Reference Format:
Anonymous Author(s). 2025. The Generative AI Divide: A Descriptive Analysis of Heterogeneous Adaptation Among Knowledge Contributors. In *Proceedings of (CSCW '26).* ACM, New York, NY, USA, 21 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

How are engagement patterns of different contributor groups within online knowledge communities changing in the context of generative AI—specifically ChatGPT? In this study, we examine behavioral shifts on Stack Overflow, one of the largest and most influential knowledge-sharing platforms where users contribute by asking questions, providing answers, and engaging with the community through activities like voting and editing [7]. Our objective is to explore the extent to which different types of contributors have changed their engagement patterns following the emergence of ChatGPT: do some groups exhibit relative resilience, or are certain groups more prone to disengagement? While prior research has documented a general decline in overall usage on Stack Overflow following the release of ChatGPT [9, 12, 13, 23, 33], it remains unclear how behavioral trends differ across user groups. Exploring these heterogeneous patterns may offer insights into the potential fragility of online knowledge sharing in the age of generative AI. Rather

CSCW '26, USA

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

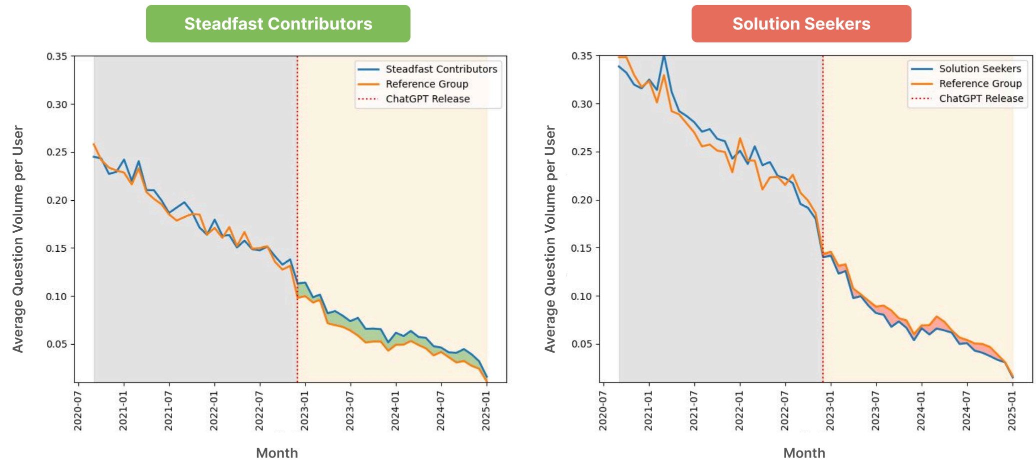

Fig. 1. Monthly question and answer volume per user for Steadfast Contributors and Solution Seekers, alongside their matched reference groups. The vertical dashed line marks the public release of ChatGPT on November 30, 2022.

than focusing solely on aggregate usage trends, we examine individual-level behavioral variation by identifying distinct contributor profiles.

To this end, we clustered 394,295 Stack Overflow users based on their pre-ChatGPT behaviors using a two-step approach: first, we applied graph-based community detection to identify groups of co-occurring badges [35]; then, we used a guided Latent Dirichlet Allocation (LDA) model to infer interpretable user roles. We focused our analysis on a subset of these users — 268,940 question askers and 259,397 answer contributors — and examined their engagement levels. For each user group, we constructed a matched reference group by calculating propensity scores based on pre-ChatGPT activity trends. We then conducted a difference-in-differences (DiD) analysis to compare relative changes in two outcome variables: the volume and length of questions and answers posted per user per month. We note that this analysis is correlational and exploratory in nature, and associations observed between generative AI and user behavior should not be interpreted as causal.

Results show that every user cluster we identified, regardless of their behavioral profile, experienced reduced participation following ChatGPT's release. While we observe some heterogeneity in the magnitude of contraction, this variation is bounded and occurs strictly within the context of platform-wide downward trend. Even the most resilient contributors reduced their participation; they simply did so less severely than their reference groups. Steadfast Contributors—highly active users embedded in the platform's infrastructure—exhibited a relatively smaller decline in both question and answer volume compared to their matched reference group. Yet critically, they still declined. Their relative resilience does not indicate immunity but rather a slower rate of disengagement from the community.

In contrast, several contributor types showed significantly steeper declines in engagement relative to their matched reference group. Peer-Conscious users—who tend to delete poorly received posts—exhibited a sharper drop in both question and answer volume. Solution Seekers—users who proactively promote and incentivize their own questions—also experienced a significantly steeper decline in both question and answer volume. Thread Revivers—users with a history of answering long-dormant questions—showed sharp reductions in both question and answer volume as well.

These findings carry several implications for understanding the fragility of online knowledge communities. First, our findings extend prior work by demonstrating that the post-ChatGPT decline is not merely aggregate but structural: it penetrates every contributor type we identified. Second, by identifying canonical user profiles and analyzing their differential trajectories, our study shows that the magnitude of decline varies across user types. Contributors with intrinsic or identity-based motivations were associated with a less severe decline, while those with instrumental or socially sensitive motivations were associated with a steeper decline in activity. Yet, this heterogeneity is bounded. The differences between contributor types are matters of degree within a uniformly negative trajectory. Finally, these patterns underscore the potential fragility of Stack Overflow's knowledge ecosystem. The declining engagement of contributors whose work is essential to sustaining archival completeness, such as Thread Revivers, points to emerging vulnerabilities in the platform's capacity to preserve the breadth and depth of its collective knowledge base. Taken together, this study sheds light on how generative AI is reshaping the online knowledge communities, highlighting how user behaviors shift and what these patterns may imply for the evolving dynamics of knowledge sharing.

In summary, this paper makes the following contributions.

- We identify bounded heterogeneity within the context of platform-wide downward trend, revealing which types of contributors exhibit comparatively milder versus sharper declines in engagement following the advent of generative AI.
- We offer perspectives on how the meaning and practice of online knowledge sharing may be evolving in the era of generative AI, showing that even the most resilient users remain subject to systemic disengagement.
- Methodologically, we present an application of topic modeling for clustering users to derive canonical user prototypes from large-scale behavioral data.

2 Related Work

2.1 User Behavior on Online Knowledge Platforms

Prior work has examined user participation dynamics in collaborative knowledge-sharing platforms like Stack Overflow [17]. User behavior on the platform reveals distinct patterns shaped by speed, expertise, and domain specialization [38]. A well-known phenomenon is the "Fastest Gun in the West" effect: early responders are more likely to be upvoted, with the median time to first answer under 16 minutes [8]. This temporal advantage reinforces rapid participation, skewing visibility and reward distribution.

While most users contribute infrequently, a small fraction of highly active users produce the bulk of content. These power users are motivated by a mix of intrinsic factors (e.g., helping others, reinforcing personal learning) and extrinsic rewards such as reputation points and badges [26, 29, 35, 42]. Long-term community success depends on fostering member commitment, and by supporting the formation of social ties, community design can shift participation from isolated interactions to enduring engagement [21, 31]. On Stack Overflow, expertise is organized around tag-based technical domains (e.g., Python, Android), forming stable communities with little cross-domain migration [28]. Specialization in a narrow range of tags is associated with higher answer quality and long-term reputation growth [25].

An unsupervised clustering of Stack Overflow users revealed additional structure [1]. In this analysis, users were grouped into behavioral archetypes — such as askers, answerers, and editors — and into reputation trajectories like naive users, surpassing users, and experts. These findings highlight that Stack Overflow operates as a collection of overlapping expert-driven subcommunities with varying motivations, skill levels, and contribution styles.

Complementary research on peer production platforms such as Wikipedia underscores the importance of behind-the-scenes activities in sustaining knowledge ecosystems. High-quality knowledge bases depend not only on visible content creation, but also on continuous, incremental tasks such as editing, cleanup, and archival maintenance [40]. These tasks often constitute forms of "invisible work" that are critical to the integrity and reliability of community knowledge [15].

2.2 Platform Adaptation to ChatGPT

Several recent studies have documented platform-wide behavioral shifts on Stack Overflow following the public release of ChatGPT [9, 12, 13, 23, 33]. One empirical analysis found a 22% decline in answer volume, along with substantial reductions in the number of questions and comments [12]. Another study reported an 11% decrease in new questions and a 12.9% decline in new user registrations [33]. These findings suggest that generative AI tools may be displacing traditional community-based knowledge exchange, particularly for routine programming inquiries.

Importantly, the decline in activity was uneven across domains [9]. Tags associated with well-documented, self-contained problems experienced sharper drops, likely because such questions are easily handled by LLMs. In contrast, tags requiring extensive context or deeper domain expertise showed more stability, pointing to differential substitution effects based on task complexity.

Del Rio-Chanona et al. [13] observed a 16% drop in weekly Stack Overflow posts immediately after ChatGPT's release, intensifying to a 25% decline by mid-2023. This trend affected both questions and answers across all user types—from first-time posters to veteran contributors. Notably, post-ChatGPT content did not suffer a detectable decline in quality, suggesting that the retreat affected high- and low-quality contributions alike. These findings raise important questions about which user segments remain engaged and how community knowledge production is evolving.

2.3 Comparing ChatGPT and Stack Overflow as Information Sources

Emerging comparisons between ChatGPT and Stack Overflow highlight user preferences and expectations. These comparisons should be viewed in the context of a rapidly evolving landscape for AI capabilities and public perception [36]. Kabir et al. [18, 19] conducted a blind comparison of AI- vs. human-generated answers to programming questions. While ChatGPT's responses were inaccurate in roughly 52% of cases, users still preferred its answers 35% of the time. Respondents cited the clarity, fluency, and comprehensiveness of ChatGPT's language — even when it was wrong — as reasons for their preference. These findings suggest that users increasingly value the form of an answer as much as, or more than, its factual correctness.

Liu et al. [24] ran a controlled experiment comparing two user groups tasked with solving programming problems — one using Stack Overflow, the other using ChatGPT. Users with AI assistance completed tasks more quickly and submitted more correct code overall. However, Stack Overflow users performed better on debugging tasks involving subtle errors, where collective human reasoning and discussion proved more effective. This suggests a task-based division of labor: users may turn to ChatGPT for boilerplate solutions and use community forums for more complex or ambiguous problems. Over time, such shifts could reshape the types of questions that are asked and answered on Stack Overflow, potentially concentrating community activity around edge cases and nuanced challenges that AI cannot yet handle.

3 Method

3.1 Data Collection

To incentivize participation, Stack Overflow awards badges—digital markers of achievement reflecting users' activity levels, expertise, and sustained engagement [35]. These badges are awarded based on measurable behaviors, such as asking or answering questions, receiving upvotes, or editing content. Importantly, badge assignment is determined algorithmically, without human moderation or subjective review. This makes badge acquisition a valuable behavioral signal for understanding user engagement over time, especially during periods of technological shifts such as the release of ChatGPT. Leveraging this infrastructure, we cluster Stack Overflow users based on their badge profiles to examine how different user groups exhibit changes in behavior around the time of generative AI.

3.1.1 Badge Selection. We began by curating a subset of 31 badges that meaningfully represent diverse user behaviors. To ensure consistency and relevance, one researcher conducted an initial review of the full list of Stack Overflow badges, and a second researcher independently reviewed the selections through iterative discussion and consensus. The final list of 31 badges is presented in Table 1. Our selection process followed three criteria:

(1) **Relevance to user behavior:** We excluded badges unrelated to knowledge contribution or community participation. For instance, the *Census* badge—awarded for completing a survey—was excluded due to its limited behavioral relevance to our research objectives.
(2) **Discriminative value:** We excluded badges that were either too common (more than 1,000,000 awardees) or too rare (fewer than 1,000 awardees), as such extremes provide limited discriminatory power for clustering. For instance, the *Student* badge—granted when a user posts their first question with a score of at least one—has been earned by over 3.2 million users and was therefore excluded due to its ubiquity. However, we retained the *Famous Question* badge as an exception: although it has been awarded to 1.2 million users, it captures behaviors highly central to our research objectives.
(3) **Redundancy reduction:** Some badges overlapped in meaning or scope. For instance, *Famous Question* ($\geq$ 10,000 views) is a strict subset of *Popular Question* ($\geq$ 1,000 views). In such cases, we retained the more general or representative badge to avoid redundancy.

3.1.2 User Sampling. We used the public Stack Exchange API v2.3 to collect user and activity data. First, using the endpoint `/badges/{ids}/recipients`, we identified all users who had earned at least one of the 31 selected badges prior to the public release of ChatGPT in November 2022 [41], resulting in a dataset of 1,277,423 users. By restricting badge acquisition to the pre-ChatGPT period, we ensured that our user clustering captures behavioral patterns that emerged independently of any influence from generative AI. This design enables a clearer examination of how distinct user groups adapted following the introduction of ChatGPT.

3.1.3 Activity Data Collection. We then collected longitudinal activity data—specifically, all questions and answers posted by these users—using the API endpoints `/users/{ids}/questions` and `/users/{ids}/answers`. To construct a balanced panel for comparison, we included only users who had posted at least one question or answer between September 1, 2020 and November 30, 2022, capturing a 27-month window before ChatGPT's release.

After filtering, the final dataset comprised 394,295 users. Of these, 268,940 users exhibited valid question activity—hereafter referred to as question askers. Similarly, 259,397 users demonstrated valid answer activity were retained—hereafter referred to as answer contributors. These datasets form the foundation for our clustering procedure and subsequent analyses.

3.2 User Clustering

To identify meaningful user groups based on their behavioral patterns, we clustered the 394,295 users who (1) had posted at least one question or answer and (2) had earned at least one of the selected 31 badges prior to the release of ChatGPT. We adopted a topic modeling approach using Latent

Table 1. Curated list of 31 Stack Overflow badges used for clustering analysis.

Badge	**Description**[a]
Inquisitive	Ask a well-received question on 30 separate days, and maintain a positive question record
Favourite Question	Question saved by 25 users
Great Question	Question score of 100 or more
Explainer	Edit and answer 1 question (both actions within 12 hours, answer score > 0)
Generalist	Provide non-wiki answers totaling a score of 15 or more in 20 of the top 40 tags
Lifejacket	Answer score of 5 or more to a question score of -2 or less, that goes on to receive a score of 2 or more
Populist	Highest scoring answer that outscored an accepted answer with a score of more than 10 by more than 2x.
Self-Learner	Answer your own question with a score of 3 or more
Tenacious	Zero score accepted answers: more than 5 and 20% of total
Pundit	Leave 10 comments with a score of 5 or more
Fanatic	Visit the site each day for 100 consecutive days (UTC)
Mortarboard	Earn at least 200 reputation (the daily maximum) in a single day
Deputy	Raise 80 helpful flags.
Reviewer	Complete at least 250 review tasks
Disciplined	Delete own post with score of 3 or higher
Strunk & White	Edit 80 posts.
Electorate	Vote on 600 questions, and 25% or more of total votes are on questions
Archaeologist	Edit 100 posts that were inactive for 6 months
Proofreader	Approve or reject 100 suggested edits
Sportsmanship	Upvote 100 answers on questions where an answer of yours has a positive score
Suffrage	Use 30 votes in a day
Altruist	First bounty you manually award on another person's question
Benefactor	First bounty you manually award on your own question
Famous Question	Question with 10,000 views
Promoter	First bounty you offer on your own question
Guru	Accepted answer and score of 40 or more
Great Answer	Answer score of 100 or more
Revival	Answer more than 30 days after a question was asked as the first answer, scoring 2 or more
Necromancer	Answer a question more than 60 days later with a score of 5 or more
Peer Pressure	Delete own post with score of -3 or lower
Favourite Answer	Answer saved by 25 users.

[a] Descriptions follow the official Stack Overflow badge documentation.

Dirichlet Allocation (LDA), framing user clustering as a role discovery task. In this formulation, users are treated as documents, badges as words, and latent behavioral profiles (or "roles") as topics. Each user's badge acquisition history is encoded as a binary vector, representing whether a badge

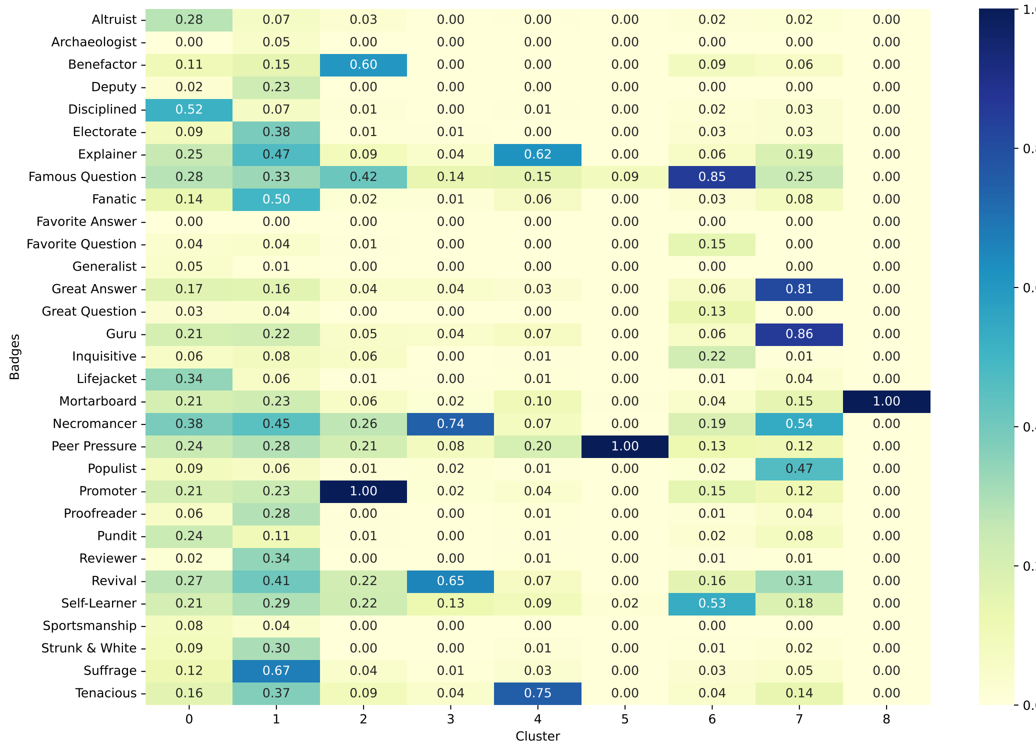

Fig. 2. Final badge clustering results using our two-step method with guided LDA. Each row represents a badge, and each column represents a user cluster. Color intensity reflects the proportion of users in a cluster who earned the badge (1.0 = all users, 0.0 = none).

was earned or not. LDA learns two key distributions from this data: (1) the user-role distribution, which indicates the probability of each user being associated with each latent role, and (2) the role-badge distribution, which captures the probability of each badge being associated with each role. To assign users to clusters, we computed the likelihood of their role membership and assigned each user to the role (cluster) with the highest probability.

3.2.1 Determining the Number of Clusters. A critical modeling decision is determining the appropriate number of clusters (topics). To address this, we empirically selected the model using a randomly sampled subset of 100,000 users. We evaluated topic quality based on two standard metrics: *coherence*, which assesses the semantic interpretability of the topics, and *perplexity*, which measures the model's fit to the data. For ease of comparison, we normalized both metrics to a [0, 1] scale—where higher coherence and lower perplexity are preferable—and computed a combined score as the average of normalized coherence and inverted perplexity. We tested models with 5 to 14 topics and found that the combined score peaked at $k = 9$, which we adopted as the optimal number of user clusters.

3.2.2 Two-Step Clustering Procedure. Our clustering procedure consisted of two steps: seed topic discovery via graph-based clustering and guided role discovery via seeded LDA.

Step 1: Graph-Based Clustering of Badges. To initialize semantically coherent seed topics for the LDA model, we first constructed a graph of badges, where each node represents a badge, each edge represents co-occurrence between two badges, and the weight of the edges represents the number of users who earned both badges. We then applied the Louvain method for community detection to partition the graph into nine groups of badges. We used these groups as seed topics for the subsequent guided LDA.

Step 2: Seeded LDA Clustering of Users. Next, we applied the guided LDA to 394,295 users, where the nine badge groups identified from graph-based clustering served as seed topics. Guided LDA biases the learning process such that specific badges are more likely to co-occur within predefined topics, thereby improving interpretability. We conducted a grid search over the hyperparameters α and β, which control the sparsity of user-role and role-badge distributions, respectively. Higher α means users having more roles, and higher β means roles having more badges.

We encountered two primary challenges: *diffusion*, where the same badge appeared across multiple clusters, and *intrusion*, where a single cluster contained too many badges. After experimenting with multiple clustering approaches [14, 22], we found that our two-step method with the guided LDA model best mitigated these issues. To further improve semantic separation, we performed a grid search over hyperparameter values, selecting those that yielded clearer distinctions between clusters. Based on manual inspection, we selected $\alpha = 0.01$ and $\beta = 0.0001$ for the model. Figure 2 presents the final clustering results, showing badge frequencies per cluster. A frequency of 1.0 indicates that all users in a given cluster earned that badge, while 0.0 indicates that none did. Table 2 reports the key badges for each cluster, defined as those with an average within—cluster frequency of at least 0.42.

Table 2. Descriptive labels for the user clusters. Each row reports the cluster's name, behavioral description, key badges, and user count.

K	**Cluster Name**	**Description**	**Key Badges (% of Users)**	**# Users**
0	Disciplined Users	These users retracted their posts, even when the posts were well-received (Disciplined). This pattern points to a high internal quality threshold that overrides external feedback.	Disciplined (52)	3,939
1	Steadfast Contributors	Marked by consistent voting (Suffrage), daily engagement (Fanatic), timely and helpful answers (Explainer), and revisiting old questions (Necromancer), these users are persistent in their contributions across many dimensions, reflecting long-term commitment and engagement.	Suffrage (67), Fanatic (50), Explainer (47), Necromancer (45)	33,073
2	Solution Seekers	These users are highly proactive in resolving their own challenges. They actively promote their questions (Promoter), often investing additional effort or incentives to attract quality responses (Benefactor). The prevalence of Famous Question suggests their posts tend to gain traction.	Promoter (100), Benefactor (60), Famous Question (42)	27,348

3	Thread Revivers	These users have a history of bringing dormant threads back to life with fresh contributions (Revival, Necromancer). Their engagement with older posts may be either intentional or incidental—encountered through search or relevance.	Necromancer (74), Revival (65)	151,904
4	Lone Solvers	These users answer questions even when their answers receive no immediate recognition (Tenacious). Their behavior may show commitment to solving problems rather than chasing votes, backing up their answers with edits and clarifications (Explainer).	Tenacious (75), Explainer (62)	12,215
5	Peer-Conscious	Defined by their willingness to retract poorly received contributions (Peer Pressure), these users exhibit high social sensitivity and a strong norm-following tendency. They likely monitor community feedback closely and self-regulate accordingly.	Peer Pressure (100)	65,639
6	Viral Askers	These users have posted one or more questions that attracted widespread attention (Famous Question).	Famous Question (85), Self-Learner (53)	95,297
7	Expert Answerers	These high-reputation users consistently provide top-quality answers (Great Answer, Guru, Populist) and specialize in reviving old questions with valuable insights (Necromancer). Their contributions often stand out in technical excellence and community impact.	Guru (86), Great Answer (81), Necromancer (54), Populist (47)	4,076
8	Peak Performers	These users hit the daily reputation ceiling (Mortarboard), suggesting bursts of highly upvoted activity. They likely contribute impactful content in short, intense periods.	Mortarboard (100)	804

3.2.3 Labeling the Clusters. To enhance the interpretability of clusters, we assigned descriptive labels to each of the nine user clusters. Table 2 presents the names and descriptions for each cluster, along with the set of key badges and the total number of users within each cluster.

To generate these labels, we employed an iterative, human-AI collaborative approach. Specifically, we prompted GPT-4o with the key badges associated with each cluster, as well as their official descriptions (see Table 1). We also provided contextual information, including an overview of our research goals and an explanation of Stack Overflow's badge system. GPT-4o was instructed to evaluate the characteristics of each badge set and propose a label and description that best captured the behavioral identity of the cluster. The outputs were evaluated based on four criteria:

(1) Clarity: The label and description should be clear and easy to understand.
(2) Distinctness: Each cluster's label should distinguish it from all others.
(3) Relevance: The label and description should be grounded in the most representative badges in the cluster.
(4) Coherence: The label and description should form a logically unified behavioral role.

One of the authors manually reviewed and iteratively refined GPT-4o's outputs to ensure that each label met the above criteria. This process resulted in a set of interpretable and behaviorally meaningful cluster labels that guide the presentation and discussion of our findings.

3.3 Analysis

To examine how different user groups changed around the time of the introduction of generative AI, we conducted a series of difference-in-differences (DiD) analyses. Specifically, for each of the nine user clusters identified via seeded LDA, we compared their behavioral change before and after the release of ChatGPT to that of a matched reference group of users. Because generative AI was introduced as a platform-wide shock, all users are exposed. Accordingly, our DiD estimates do not identify a causal effect of AI adoption. Instead, they capture relative deviations in post-ChatGPT activity trajectories between a focal cluster and a behaviorally similar reference group.

3.3.1 Matched Reference Group Construction. The public release of ChatGPT constituted a platform-wide shock affecting all users. Thus, our analysis does not rely on an untreated control group. Instead, for each cluster, we construct a matched reference group consisting of users from other clusters who exhibited highly similar behavioral trajectories prior to ChatGPT's release. This procedure enables a valid DiD comparison to capture relative divergence in post-ChatGPT activity across clusters [6]. For each of the nine clusters, we applied the following steps.

Step 1: Identifying focal users. For a given cluster, we defined the focal group as the set of users belonging to that cluster. The matched reference group was drawn from users in the remaining clusters.

Step 2: Constructing per-user features. For all users, we computed a set of summary statistics based on their activity during the pre-ChatGPT period. Specifically, for each user, we calculated the mean, standard deviation, and linear trend (slope) of two key metrics: (1) the volume of contributions (i.e., number of questions or answers per month), and (2) the total length of their contributions in characters. We also recorded the number of active months. These seven variables constituted the matching covariates.

Step 3: Estimating propensity scores. We estimated propensity scores using logistic regression, predicting the likelihood that a user belonged to the focal cluster based on the standardized covariates. To ensure overlap in pre-period behavior, we applied a common support condition by restricting the donor pool to users whose propensity scores fell within a narrow caliper (± 0.05) of the focal group's score range.

Step 4: One-to-one matching without replacement. Within the reduced donor pool, we paired each focal user with exactly one reference user. By default, matches were selected using Mahalanobis distance computed over the covariates, which minimizes multivariate distance while accounting for their covariate structure. For only one cluster ($K = 3$) with more than 120,000 focal users, constructing a full distance matrix was pointless, so we instead employed greedy nearest-neighbor matching based on the propensity scores. The resulting matched samples contain equal numbers of focal and reference users.

These procedures were designed to address common limitations of relying solely on propensity score matching [20]. Standardizing covariates prior to estimation helps mitigate bias arising from differences in measurement scales. Sampling without replacement preserves heterogeneity within the reference pool and prevents over-representation of individual users. In addition, combining Mahalanobis distance with a caliper-restricted common support region reduces multivariate imbalance more effectively than propensity-score-only matching, thereby enhancing the internal validity of the matched reference groups. Table 3 shows the standardized difference of the covariates before matching, computed by comparing the focal users to all other users, and after matching, computed by comparing the focal users to their matched reference users. All previously unacceptable imbalances, defined as absolute standardized differences greater than 0.1, have been successfully reduced to acceptable levels (i.e., below 0.1), in accordance with the threshold suggested by Austin et al. [4, 5, 16].

Table 3. Standardized mean differences in pre-period questions and answers between focal and comparison groups. “Before” compares focal users to all non-focal users; “After” compares focal users to their matched reference users only.

		Question Askers		**Answer Contributors**	
K	Cluster	Before	After	Before	After
0	Disciplined Users	0.016	0.002	0.151	0.003
1	Steadfast Contributors	0.028	0.002	0.201	0.048
2	Solution Seekers	0.130	0.005	-0.026	0.004
3	Thread Revivers	-0.141	0.001	-0.110	-0.041
4	Lone Solvers	-0.057	0.001	0.219	0.004
5	Peer-Conscious	-0.060	-0.001	-0.108	0.002
6	Viral Askers	0.091	0.022	-0.094	0.004
7	Expert Answerers	-0.160	0.000	-0.034	0.003
8	Peak Performers	-0.008	0.002	0.034	0.003

3.3.2 Difference-in-Differences. To evaluate whether the post-ChatGPT engagement trajectories of different user clusters diverged from those of behaviorally similar peers, we employed a difference-in-differences (DiD) analysis. For each user cluster, we compared changes in behavior before and after the public release of ChatGPT with changes observed in a matched reference group using the following regression specification:

$$E[Y_{it}] = \exp\left(\beta_1 \cdot \text{Post}_t + \beta_2 \cdot (\text{Focal}_i \times \text{Post}_t) + \alpha_i + \delta_m\right) \tag{1}$$

Here, Y_{it} is the outcome variable for user i at time t, $\texttt{Focal}_i$ is a binary indicator for whether the user belongs to the focal cluster under analysis, and $\texttt{Post}_t$ is a binary indicator for whether the observation occurs after the public release of ChatGPT (November 30, 2022). The model includes user fixed effects (α_i) to control for time-invariant individual heterogeneity and month-of-year fixed effects (δ_m) to account for seasonality. Note that the main effect of $\texttt{Focal}_i$ is absorbed by the user fixed effects and thus does not appear separately in the equation.

The coefficient β_1 captures the change in the log expected outcome for the reference group in the post-period relative to the pre-period. The coefficient β_2 on the interaction term $\texttt{Focal}_i \times \text{Post}_t$ captures the relative difference in the post-period proportional change between the focal cluster and the matched reference group. Specifically, $\exp(\beta_2)$ represents the ratio of the focal group's post-to-pre change relative to the reference group's post-to-pre change. Because all users are exposed to the introduction of generative AI, this coefficient should be interpreted as a descriptive measure of relative behavioral change rather than a causal treatment effect.

We conducted this analysis using two outcome variables for both questions and answers:

- **Volume:** the average number of questions or answers posted per user per month.
- **Length:** the average length of questions or answers in characters.

Each regression model was estimated using Pseudo-Poisson Maximum Likelihood (PPML) with two-way fixed effects. The PPML estimator is particularly suited to our data, which is strictly non-negative, characterized by a high proportion of zeros (93.46%) and a dispersion index of 2.83. Unlike OLS with log-transformed outcomes, PPML remains consistent under these conditions and avoids the bias introduced by log transformations of zero-valued observations. Standard errors were clustered at the user level to account for serial correlation in user activity over time.

We fit each model separately for each user cluster and its matched reference group. To address multiple hypothesis testing across the nine clusters, we applied the Benjamini-Hochberg procedure to control the false discovery rate (FDR). For each model, we report the estimated coefficient β_2, its statistical significance after FDR correction, and confidence intervals to assess the extent to which the post-ChatGPT activity trajectory of a given cluster diverges from that of behaviorally similar peers.

Our analysis retains all users in the panel and codes post-ChatGPT months with zero activity rather than dropping them, to avoid conditioning on post-period behavior. Consequently, our estimates reflect changes in overall engagement at the cluster level, incorporating both exits and changes among still-active users, so any average reductions from departures are part of the effect, not an artifact.

Table 4. Difference in pre-ChatGPT slopes between the focal user group and its matched reference group.

		Question Askers			Answer Contributors		
K	Cluster	# Users	γ_2	p	# Users	γ_2	p
0	Disciplined Users	2,678	-1.461×10^{-4}	0.971	2,683	-2.106×10^{-4}	0.972
1	Steadfast Contributors	21,479	2.324×10^{-4}	0.874	25,543	6.390×10^{-4}	0.765
2	Solution Seekers	24,329	2.394×10^{-4}	0.855	15,688	7.061×10^{-4}	0.821
3	Thread Revivers	68,604	7.518×10^{-5}	0.917	126,919	-4.520×10^{-4}	0.612
4	Lone Solvers	6,556	1.169×10^{-4}	0.963	10,072	3.266×10^{-4}	0.910
5	Peer-Conscious	59,144	7.815×10^{-4}	0.431	25,058	1.890×10^{-3}	0.288
6	Viral Askers	83,949	-1.686×10^{-4}	0.807	49,527	7.919×10^{-4}	0.623
7	Expert Answerers	1,761	7.754×10^{-4}	0.865	3,273	3.979×10^{-4}	0.948
8	Peak Performers	440	5.764×10^{-4}	0.961	634	1.812×10^{-3}	0.868

3.4 Parallel Trend Assumption

Because our analysis is descriptive, we use DiD as a tool to characterize relative changes over time across groups. A standard diagnostic in this setting is whether groups display similar outcome trajectories prior to the focal event, which helps ensure that post-period divergences are not driven by pre-existing trend differences [6].

We describe these patterns by estimating the following model restricted to the pre-ChatGPT period.

$$E[Y_{it}] = \exp\left(\gamma_1 \cdot t + \gamma_2 \cdot (t \times \text{Focal}_i) + \alpha_i + \delta_m\right) \tag{2}$$

In this specification, Y_{it} denotes user i's average monthly questions or answers at time t, and $\texttt{Focal}_i$ indicates whether the user i belongs to the focal cluster under analysis. The variable t is the relative time (in months) leading up to the release of ChatGPT. The model includes user fixed effects (α_i) to control for time-invariant individual heterogeneity and month-of-year fixed effects (δ_m) to account for seasonality. Here, the coefficient γ_1 captures the linear pre-ChatGPT temporal trend for the reference group. The coefficient of interest, γ_2, represents the difference in the proportional trend between the focal cluster and the matched reference group.

Table 4 reports the estimated γ_2 coefficients for each cluster. For both questions and answers, we fail to reject the null hypothesis of equal pre-trends for all groups ($p > 0.05$ in all cases), indicating no detectable divergence in pre-ChatGPT slopes between focal and reference groups.

4 Results

4.1 Question Volume

Table 5 shows the regression results for the average question volume per month.

Table 5. Difference-in-Differences estimates for average monthly questions

Cluster	β_1		β_2			Effect size (%)
	Estimate	95% CI	Estimate	95% CI	p	
Disciplined Users	-1.191^{***}	$[-1.309, -1.073]$	-0.083	$[-0.240, 0.074]$	0.300	−8.0
Steadfast Contributors	-1.238^{***}	$[-1.274, -1.202]$	0.168^{***}	$[0.119, 0.217]$	<0.001	+18.3
Solution Seekers	-1.224^{***}	$[-1.259, -1.190]$	-0.105^{***}	$[-0.151, -0.059]$	<0.001	−10.0
Thread Revivers	-1.151^{***}	$[-1.170, -1.132]$	-0.056^{***}	$[-0.083, -0.029]$	<0.001	−5.4
Lone Solvers	-1.149^{***}	$[-1.210, -1.089]$	-0.057	$[-0.146, 0.031]$	0.227	−5.6
Peer-Conscious	-1.174^{***}	$[-1.196, -1.153]$	-0.591^{***}	$[-0.627, -0.555]$	<0.001	−44.6
Viral Askers	-1.339^{***}	$[-1.357, -1.321]$	0.245^{***}	$[0.220, 0.270]$	<0.001	+27.8
Expert Answerers	-1.135^{***}	$[-1.254, -1.016]$	0.105	$[-0.055, 0.265]$	0.227	+11.1
Peak Performers	-1.455^{***}	$[-1.655, -1.254]$	-0.257	$[-0.560, 0.046]$	0.145	−22.7

Notes: β_1 captures the pre/post change for the matched reference group; β_2 is the DiD estimate (Focal × Post interaction). Effect size = $(\exp(\beta_2) - 1) \times 100$, expressed as a percentage. p-values are FDR-adjusted using the Benjamini-Hochberg procedure. Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

4.1.1 Universal Decline. The β_1 estimate is negative in every cluster, ranging from −1.135 to −1.455 (all $p < 0.001$). This indicates that all reference groups reduced their question-asking after ChatGPT's release. These results are consistent with prior work [9, 13, 23].

4.1.2 Cluster-specific Divergence. While all users reduced their question volume, we observe modest variation in the severity of the decline. Two clusters exhibited a less severe decrease: Viral Askers ($\beta_2 = +0.245$, $p < 0.001$) and Steadfast Contributors ($\beta_2 = +0.168$, $p < 0.001$) showed post-to-pre ratios 27.8% and 18.3% higher then their matched reference groups, respectively. However, it is crucial to note that these users still experienced large absolute declines—they simply fell somewhat less sharply than comparable users.

By contrast, Peer-Conscious users experienced the most severe disproportionate contraction ($\beta_2 = -0.591$, $p < 0.001$), with their post-to-pre ratio 44.6% lower than their matched reference group. Solution Seekers ($\beta_2 = -0.105$, $p < 0.001$) and Thread Revivers ($\beta_2 = -0.056$, $p < 0.001$) also experienced larger contractions, with post-to-pre ratios 10.0% and 5.4% lower, respectively, relative to their matched reference group.

For the remaining clusters—Disciplined Users, Lone Solvers, Expert Answerers, and Peak Performers—the interaction coefficients were not statistically significant ($p > 0.05$), implying no detectable divergence in their post-ChatGPT trajectories relative to their matched reference group.

4.2 Question Length

Table 6 presents the regression results for question length in characters.

4.2.1 Platform-wide Shift. Despite the overall decline in question volume, reference group users for eight of the nine clusters exhibited statistically significant increases in question length after ChatGPT. On the log scale, these pre-post shifts in the reference group (β_1) ranged from +0.071 for Lone Solvers to +0.128 for Steadfast Contributors, corresponding to approximately 7.4% to 13.6% increase in mean length. The only exception was Peak Performers, whose reference group showed no significant change ($\beta_1 = -0.002,\ p = 0.975$).

4.2.2 Cluster-specific Divergence. Only one cluster showed a statistically significant divergence from its matched reference group. Peer-Conscious users exhibited a steeper increase in question length ($\beta_2 = 0.074,\ p < 0.001$), corresponding to approximately 7.6% higher post-to-pre ratio relative to their matched reference group in the post-ChatGPT period.

For all other clusters, the interaction coefficients were not statistically significant ($p > 0.05$), implying no detectable divergence in their post-ChatGPT question length trajectories relative to their matched reference group.

Table 6. Difference-in-Differences estimates for average question length (characters)

Cluster	β_1		β_2			Effect size (%)
	Estimate	95% CI	Estimate	95% CI	p	
Disciplined Users	0.115^{***}	[0.056, 0.174]	−0.043	[−0.129, 0.042]	0.426	−4.2
Steadfast Contributors	0.128^{***}	[0.109, 0.147]	−0.030	[−0.056, −0.003]	0.082	−2.9
Solution Seekers	0.084^{***}	[0.068, 0.100]	−0.006	[−0.028, 0.016]	0.587	−0.6
Thread Revivers	0.123^{***}	[0.108, 0.137]	−0.025	[−0.044, −0.005]	0.063	−2.4
Lone Solvers	0.071^{***}	[0.033, 0.109]	0.040	[−0.016, 0.096]	0.300	+4.1
Peer-Conscious	0.118^{***}	[0.104, 0.132]	0.074^{***}	[0.051, 0.096]	< 0.001	+7.6
Viral Askers	0.123^{***}	[0.113, 0.133]	−0.011	[−0.024, 0.002]	0.234	−1.1
Expert Answerers	0.114^{*}	[0.012, 0.216]	−0.038	[−0.169, 0.093]	0.587	−3.7
Peak Performers	−0.002	[−0.155, 0.150]	0.101	[−0.103, 0.306]	0.426	+10.7

Notes: β_1 captures the pre/post change for the matched reference group; β_2 is the DiD estimate (Focal × Post interaction). Effect size = $(\exp(\beta_2) - 1) \times 100$, expressed as a percentage. p-values are FDR-adjusted using the Benjamini-Hochberg procedure. Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

4.3 Answer Volume

Table 7 shows the regression results for the average answer volume per month.

4.3.1 Universal Decline. Every reference group saw a statistically significant drop in answer volume (all $\beta_1 < 0, p < 0.001$). These results are consistent with prior work [9, 13, 23].

4.3.2 Cluster-specific Divergence. Three clusters exhibited a less severe decline than their matched reference group. Expert Answerers showed relative resilience ($\beta_2 = 0.289,\ p < 0.05$), meaning their post-to-pre ratio was 33.5% higher relative to their matched reference group. Disciplined Users ($\beta_2 = 0.269,\ p < 0.05$) and Steadfast Contributors ($\beta_2 = 0.218,\ p < 0.001$) also demonstrated relative resilience, with approximately 30.8% and 24.3% higher post-to-pre ratio relative to their matched reference group, respectively.

By contrast, Peer-Conscious users experienced the largest disproportionate contraction ($\beta_2 = -0.442,\ p < 0.001$), representing approximately 35.7% lower post-to-pre ratio relative to their matched reference group. Solution Seekers ($\beta_2 = -0.181,\ p < 0.05$), Lone Solvers ($\beta_2 = -0.151,\ p < 0.05$), Viral Askers ($\beta_2 = -0.140,\ p < 0.001$), and Thread Revivers ($\beta_2 = -0.129,\ p < 0.001$) also experienced significantly larger contractions, with approximately 16.6%, 14.0%, 13.1%, and 12.1% lower post-to-pre ratio, respectively, relative to their matched reference group.

For Peak Performers, the interaction coefficient is not statistically significant ($p = 0.683$), implying no detectable divergence in their post-ChatGPT trajectories relative to their matched reference group.

Table 7. Difference-in-Differences estimates for average monthly answers

Cluster	β_1		β_2			Effect size (%)
	Estimate	95% CI	Estimate	95% CI	p	
Disciplined Users	-1.083^{***}	$[-1.218,\ -0.948]$	0.269^{*}	$[0.051,\ 0.487]$	0.021	+30.8
Steadfast Contributors	-1.278^{***}	$[-1.334,\ -1.222]$	0.218^{***}	$[0.135,\ 0.301]$	<0.001	+24.3
Solution Seekers	-1.207^{***}	$[-1.333,\ -1.081]$	-0.181^{*}	$[-0.329,\ -0.033]$	0.021	−16.6
Thread Revivers	-1.061^{***}	$[-1.114,\ -1.008]$	-0.129^{***}	$[-0.189,\ -0.070]$	<0.001	−12.1
Lone Solvers	-1.257^{***}	$[-1.348,\ -1.167]$	-0.151^{*}	$[-0.284,\ -0.018]$	0.030	−14.0
Peer-Conscious	-1.176^{***}	$[-1.256,\ -1.097]$	-0.442^{***}	$[-0.540,\ -0.344]$	<0.001	−35.7
Viral Askers	-1.009^{***}	$[-1.067,\ -0.951]$	-0.140^{***}	$[-0.210,\ -0.070]$	<0.001	−13.1
Expert Answerers	-1.094^{***}	$[-1.231,\ -0.956]$	0.289^{*}	$[0.072,\ 0.505]$	0.016	+33.5
Peak Performers	-1.688^{***}	$[-2.036,\ -1.339]$	0.116	$[-0.441,\ 0.674]$	0.683	+12.3

Notes: β_1 captures the pre/post change for the matched reference group; β_2 is the DiD estimate (Focal × Post interaction). Effect size = $(\exp(\beta_2) - 1) \times 100$, expressed as a percentage. p-values are FDR-adjusted using the Benjamini-Hochberg procedure. Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

4.4 Answer Length

Table 8 presents the regression results for answer length in characters.

4.4.1 Platform-wide Shift. The reference groups for seven of the nine clusters exhibited statistically significant increases in answer length after ChatGPT. On the log scale, these pre-post shifts in the reference group (β_1) ranged from +0.039 for Solution Seekers to +0.133 for Peer-Conscious users, corresponding to approximately 4.0% to 14.2% increases in mean length. The exceptions were Expert Answerers ($p = 0.055$) and Peak Performers ($p = 0.418$), whose reference group showed no significant change.

4.4.2 Cluster-specific Divergence. Only one cluster showed a statistically significant divergence from its matched reference group. Thread Revivers exhibited a significantly smaller increase in answer length compared to their matched reference group ($\beta_2 = -0.018,\ p < 0.05$), corresponding to approximately 1.8% lower post-to-pre ratio relative to their matched reference group in the post-ChatGPT period.

Table 8. Difference-in-Differences estimates for average answer length (characters)

Cluster	β_1 Estimate	β_1 95% CI	β_2 Estimate	β_2 95% CI	p	Effect size (%)
Disciplined Users	0.049^{**}	$[0.015, 0.083]$	0.028	$[-0.019, 0.075]$	0.447	+2.8
Steadfast Contributors	0.060^{***}	$[0.048, 0.071]$	0.013	$[-0.003, 0.029]$	0.356	+1.3
Solution Seekers	0.039^{**}	$[0.015, 0.063]$	0.022	$[-0.012, 0.056]$	0.447	+2.2
Thread Revivers	0.052^{***}	$[0.043, 0.060]$	-0.018^{*}	$[-0.030, -0.006]$	0.028	−1.8
Lone Solvers	0.085^{***}	$[0.068, 0.101]$	0.012	$[-0.014, 0.037]$	0.550	+1.2
Peer-Conscious	0.133^{***}	$[0.107, 0.159]$	−0.003	$[-0.043, 0.037]$	0.897	−0.3
Viral Askers	0.052^{***}	$[0.038, 0.067]$	−0.023	$[-0.044, -0.002]$	0.137	−2.3
Expert Answerers	0.048	$[-0.001, 0.098]$	−0.017	$[-0.082, 0.048]$	0.680	−1.7
Peak Performers	0.042	$[-0.060, 0.144]$	−0.052	$[-0.220, 0.116]$	0.680	−5.1

Notes: β_1 captures the pre/post change for the matched reference group; β_2 is the DiD estimate (Focal × Post interaction). Effect size = $(\exp(\beta_2) - 1) \times 100$, expressed as a percentage. p-values are FDR-adjusted using the Benjamini-Hochberg procedure. Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

For all other clusters, the interaction coefficients are not statistically significant ($p > 0.05$), implying no detectable divergence in their post-ChatGPT answer length trajectories relative to their matched reference group.

Table 9. Summary of our results in terms of DiD effect sizes

Cluster	Description	Question Volume	Question Length	Answer Volume	Answer Length
Disciplined Users	Users who deleted posts despite positive feedback.	—	—	+30.8 %	—
Steadfast Contributors	Consistently active and deeply engaged across various activities.	+18.3 %	—	+24.3 %	—
Solution Seekers	Promoted and enhanced their questions to get better answers.	-10.0 %	—	-16.6 %	—
Thread Revivers	Re-engaged with and updated old, dormant threads.	-5.4 %	—	-12.1 %	-1.8 %
Lone Solvers	Persistently answered questions without needing recognition.	—	—	-14.0 %	—
Peer-Conscious	Remove poorly received posts to align with community norms.	-44.6 %	+7.6 %	-35.7 %	—
Viral Askers	Their questions gained widespread attention.	+27.8 %	—	-13.1 %	—
Expert Answerers	Provided consistently top-quality, technical answers.	—	—	+33.5 %	—
Peak Performers	Contributed high-impact content in intense bursts.	—	—	—	—

Notes: Effect sizes are reported only for statistically significant coefficients. Green indicates a less severe decline or a larger increase, while red indicates a more severe decline or a smaller increase relative to the matched reference group.

5 Discussion

Our analysis shows that the introduction of ChatGPT coincided with a broad contraction in Stack Overflow engagement. Every user cluster, regardless of their behavioral profile, experienced a decline. While we observe some heterogeneity in the magnitude of contraction, this variation occurs strictly within the context of a platform-wide downward trend. Even the most resilient contributors reduced their participation; they simply did so less severely than their reference groups. This pattern suggests a degree of fragility in Stack Overflow's knowledge-sharing network in the era of generative AI.

5.1 Who remains resilient in the face of AI?

While overall engagement declined following ChatGPT's release, some contributor groups demonstrated resilience. The resilience here refers to a comparatively smaller reduction in contribution behavior following the external technological shift. In particular, Steadfast Contributors—highly active contributors embedded in the norms and infrastructure of the platform—showed relatively smaller declines in both question and answer volume compared to their matched reference group. Viral askers experienced a comparatively milder reduction in question volume than their matched reference group, while Disciplined Users and Expert Answerers showed somewhat less severe drops in answer volume relative to their matched reference groups.

Steadfast Contributors exhibited a post-to-pre ratio of question volume that was 18.3% higher than that of their matched reference group. They also showed a post-to-pre ratio of answer volume that was 24.3% higher than their matched reference group. Prior research on online communities highlights the importance of social identity in sustaining long-term engagement. According to social identity theory, users who identify strongly with a community's values are more likely to maintain participation over time [21, 30, 31]. In the case of Steadfast Contributors, motivation may stem not only from obtaining information but from identity-based and intrinsic drivers, such as contributing to a shared resource or reinforcing one's social role [39]. However, it is crucial to contextualize these findings within the broader platform collapse: even Steadfast Contributors experienced absolute declines in activity; they simply declined *less severely* than their matched reference group. This limited heterogeneity, where the best case scenario is a slower rate of decline rather than stability or growth, reveals a potential network fragility of Stack Overflow. When even the most committed users cannot escape the downward trajectory, it suggests that generative AI's disruption operates at a systemic level that individual user resilience cannot counteract. The heterogeneity we observe is not a sign of platform health, but is better understood as variation in the pace of disengagement from a community experiencing a broad contraction.

5.2 Who is more likely to disengage from the community?

In contrast, several contributor groups exhibited steeper declines in participation following ChatGPT's emergence.

Peer-Conscious users—those attuned to social evaluation and quick to delete poorly received posts—exhibited a marked reduction in both question and answer volume, significantly greater than that of their matched reference group. This behavioral shift may reflect heightened sensitivity to social evaluation, consistent with theories of social conformity [2, 3] and performance pressure [10]. The presence of a public audience capable of judging one's contributions is known to increase anxiety and may impair or inhibit action. This theoretical foundation helps explain why Peer-Conscious users may stop posting questions and answers: the fear of disapproval from the community discourages public posting.

Solution Seekers exhibited a pronounced decline in both question and answer volume—more severe than that of their matched reference group. These users tend to be proactive in soliciting help, promoting their own queries, and incentivizing responses. This decline coincided with the rise of generative AI tools, which could be associated with a shift in how these users seek assistance. For this group, tools like ChatGPT offer a compelling value proposition: instant feedback without the delay, uncertainty, or social friction inherent in community-based platforms. This shift may be interpreted through the lens of the Google Effect [34], which posits that individuals tend to offload cognitive tasks to readily accessible technologies. Applying this lens, Solution Seekers may now bypass the community in favor of AI tools that satisfy their instrumental need, as their motivation appears rooted not in social interaction or knowledge sharing, but in outcome-oriented problem resolution.

Thread Revivers have a history of responding to archived or long-dormant questions. This activity may be either intentional, driven by interest in unresolved topics, or incidental, arising from encounters via search or exploration. This group experienced a steeper decline in the volume of questions and answers, significantly greater than that of their matched reference group. One plausible explanation is that their maintenance-oriented contributions tend to be low in visibility and receive limited social recognition, which may be associated with diminished motivation over time [32, 37]. As AI tools increasingly fulfill informational needs, the perceived impact of such contributions may decline. Alternatively, if Thread Revivers are primarily motivated by extrinsic goals rather than a sense of community identification, they may be vulnerable to disengagement—similar to Solution Seekers or Peer-Conscious users.

5.3 Implications

This study prompts several considerations for the future of online knowledge communities.

First, our results extend prior work documenting a platform-wide decline in user engagement following the release of ChatGPT [9, 12, 13, 23]. Consistent with this literature, we find a broad contraction in both question and answer activity, reconfirming that the post-ChatGPT period is characterized by a system-wide reduction in participation.

Second, we contribute a methodological approach to this domain by identifying canonical user prototypes through unsupervised clustering based on behavioral data. Our analysis shows that while the magnitude of decline varies across user types, this heterogeneity is bounded. While groups such as Peer-Conscious users, Solution Seekers, and Thread Revivers experience a steeper drop in activity, Steadfast Contributors exhibit a less severe decline. Importantly, the variation we observe represents differences in the rate of disengagement, not in its direction: all groups ultimately follow the same downward trajectory.

Lastly, the fragility exposed by this disengagement trend raises questions about the long-term sustainability of community-based knowledge platforms. For instance, the decline in engagement among Thread Revivers raises concerns about the potential erosion of socially and structurally important roles. These users have historically contributed by maintaining the long-tail content of the platform—reviving unresolved threads and enhancing archival completeness. Their reduced activity suggests the possibility that generative AI tools may be associated with lower participation from contributors whose work, while often low in visibility, plays a critical role in maintaining the integrity of the knowledge base. If such users disengage, the platform may experience a gradual loss in breadth and archival depth, potentially narrowing the scope of knowledge sharing over time.

5.4 Limitations

Nevertheless, our study has limitations. First, there is potential selection bias due to our sampling strategy. We focused on users who earned selected badges and had contributed during the pre-ChatGPT period, which skews the sample toward more active users. However, this is intentional and appropriate given our research focus on behavioral adaptation among active contributors, who are known to generate the majority of content on Stack Overflow. Our findings do not generalize to all users, as we specifically target particular contributor groups, such as Steadfast Contributors or Thread Revivers. Because both clustered users and their matched reference group are badge-earning users, our findings reflect behavioral patterns within the subpopulation of actual contributors rather than the average Stack Overflow user.

Second, our dataset does not permit direct inference about individuals' underlying motivations. Accordingly, we acknowledge that some of our role labels and interpretations may be read as extending beyond what is strictly observed in the data. All such interpretive labels are explicitly characterized as speculative and heuristic in nature, intended to facilitate discussion rather than to advance claims about latent motivational states. This constitutes a limitation of our study: motivational interpretations are not directly identified and should therefore be treated with appropriate caution.

Third, our interpretations of behavioral change are inferential rather than grounded in direct user feedback. For example, we inferred that longer question length may be associated with more complex problems. However, it is also plausible that the increase is related to users copying and pasting ChatGPT-generated content into their posts. Other unobserved motivations may also be at play. Future research should complement this analysis with qualitative studies or user interviews to more precisely capture the intent behind these behavioral shifts.

Finally, we note that this is a descriptive study. We rely on archival platform data and apply observational study methods to characterize heterogeneous adaptation among knowledge contributors. While difference-in-differences modeling helps identify divergences in trajectories, the analysis remains correlational and does not establish causal relationships. Our contribution is therefore best understood as documenting descriptive patterns of adaptation across contributor groups in the wake of generative AI, rather than identifying definitive causal mechanisms.

6 Conclusion

This study investigates how different types of contributors on online knowledge communities adapted to the emergence of generative AI, focusing on engagement before and after the public release of ChatGPT. By clustering Stack Overflow users into interpretable behavioral profiles and analyzing their engagement patterns through a difference-in-differences analysis, we find that the platform experienced a broad contraction in activity—yet within this decline, bounded heterogeneity exists across user roles. Even the most resilient groups, such as Steadfast Contributors, showed declines in question and answer volume, albeit milder than their matched reference group. Contributors such as Peer-Conscious users, Solution Seekers, and Thread Revivers exhibited significantly steeper drops in engagement compared to their matched reference groups.

These findings reveal that generative AI has precipitated a platform-wide downward trend from which no contributor type has been immune. While heterogeneity in responses exists, it is constrained: the differences lie not in whether users disengaged, but in how sharply. Contributors with intrinsic or identity-based motivations may be comparatively more resilient, whereas those driven by instrumental goals or socially sensitive dynamics may be more susceptible to disengagement. The universal nature of this decline, including among structurally important roles, raises concerns about the long-term sustainability of online knowledge communities.

Taken together, our findings offer a nuanced view of how engagement patterns are shifting within knowledge-sharing communities in the age of generative AI. We identify which types of contributors exhibit comparatively milder versus sharper declines in engagement following the advent of generative AI. Rather than assessing only aggregate changes, this study aims to understand the bounded heterogeneity within the context of platform-wide collapse and what these patterns may imply for the future of knowledge-sharing platforms.

References

[1] J Anusha, V Smrithi Rekha, and P Bagavathi Sivakumar. 2015. A machine learning approach to cluster the users of stack overflow forum. In *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems: Proceedings of ICAEES 2014, Volume 2*. Springer, 411–418.

[2] Solomon E Asch. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied* 70, 9 (1956), 1.

[3] Solomon E Asch. 2016. Effects of group pressure upon the modification and distortion of judgments. In *Organizational influence processes*. Routledge, 295–303.

[4] Peter C Austin. 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine* 28, 25 (2009), 3083–3107.

[5] Peter C Austin, Paul Grootendorst, and Geoffrey M Anderson. 2007. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Statistics in medicine* 26, 4 (2007), 734–753.

[6] Andrew Baker, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. 2025. Difference-in-Differences Designs: A Practitioner's Guide. *arXiv preprint arXiv:2503.13323* (2025).

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

[8] Vasudev Bhat, Adheesh Gokhale, Ravi Jadhav, Jagat Pudipeddi, and Leman Akoglu. 2014. Min (e) d your tags: Analysis of question response time in stackoverflow. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*. ieee, 328–335.

[9] Gordon Burtch, Dokyun Lee, and Zhichen Chen. 2024. The consequences of generative AI for online knowledge communities. *Scientific Reports* 14, 1 (2024), 10413.

[10] Nickolas B Cottrell. 1972. Social facilitation. *Experimental social psychology* 185 (1972), 236.

[11] Carla Curado and Nick Bontis. 2006. The knowledge-based view of the firm and its theoretical precursor. *International Journal of Learning and Intellectual Capital* 3, 4 (2006), 367–381.

[12] Leuson Da Silva, Jordan Samhi, and Foutse Khomh. 2024. Chatgpt vs llama: Impact, reliability, and challenges in stack overflow discussions. *arXiv preprint arXiv:2402.08801* (2024).

[13] R Maria del Rio-Chanona, Nadzeya Laurentsyeva, and Johannes Wachs. 2024. Large language models reduce public knowledge sharing on online Q&A platforms. *PNAS nexus* 3, 9 (2024), pgae400.

[14] K-L Du. 2010. Clustering: A neural network approach. *Neural networks* 23, 1 (2010), 89–107.

[15] R Stuart Geiger and David Ribes. 2010. The work of sustaining order in Wikipedia: The banning of a vandal. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 117–126.

[16] Kartik Hosanagar, Daniel Fleder, Dokyun Lee, and Andreas Buja. 2014. Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science* 60, 4 (2014), 805–823.

[17] Elina H Hwang and David Krackhardt. 2020. Online knowledge communities: breaking or sustaining knowledge silos? *Production and Operations Management* 29, 1 (2020), 138–155.

[18] Samia Kabir, David N Udo-Imeh, Bonan Kou, and Tianyi Zhang. 2023. Who answers it better? an in-depth analysis of chatgpt and stack overflow answers to software engineering questions. *CoRR* (2023).

[19] Samia Kabir, David N Udo-Imeh, Bonan Kou, and Tianyi Zhang. 2024. Is stack overflow obsolete? an empirical study of the characteristics of chatgpt answers to stack overflow questions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[20] Gary King and Richard Nielsen. 2019. Why Propensity Scores Should Not Be Used for Matching. *Political Analysis* 27, 4 (2019 2019), 435–454. https://doi.org/10.1017/pan.2019.11

[21] Robert E Kraut and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design*. Mit Press.

[22] Dokyun "DK" Lee, Zhaoqi "ZQ" Cheng, Chengfeng Mao, and Emaad Manzoor. 2025. Guided diverse concept miner (GDCM): Uncovering relevant constructs for managerial insights from text. *Information systems research* 36, 1 (2025), 370–393.

[23] Xinyu Li and Keongtae Kim. 2024. Impacts of generative AI on user contributions: evidence from a coding Q &A platform. *Marketing Letters* (2024), 1–15.

[24] Jinrun Liu, Xinyu Tang, Linlin Li, Panpan Chen, and Yepang Liu. 2023. 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)*. IEEE, 364–373.

[25] Giacomo Livan, Giuseppe Pappalardo, and Rosario N Mantegna. 2022. Quantifying the relationship between specialisation and reputation in an online platform. *Scientific Reports* 12, 1 (2022), 16699.

[26] Yao Lu, Xinjun Mao, Minghui Zhou, Yang Zhang, Zude Li, Tao Wang, Gang Yin, and Huaimin Wang. 2021. Motivation under gamification: An empirical study of developers' motivations and contributions in stack overflow. *IEEE transactions on software engineering* 48, 12 (2021), 4947–4963.

[27] Thomas W Malone, Robert Laubacher, and Chrysanthos Dellarocas. 2010. The collective intelligence genome. *MIT Sloan management review* (2010).

[28] Iraklis Moutidis and Hywel TP Williams. 2021. Community evolution on stack overflow. *Plos one* 16, 6 (2021), e0253010.

[29] Oded Nov. 2007. What motivates wikipedians? *Commun. ACM* 50, 11 (2007), 60–64.

[30] Yuqing Ren, Robert Kraut, and Sara Kiesler. 2007. Applying common identity and bond theory to design of online communities. *Organization studies* 28, 3 (2007), 377–408.

[31] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. 1994. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. 175–186.

[32] Michael Restivo and Arnout Van De Rijt. 2012. Experimental study of informal rewards in peer production. *PloS one* 7, 3 (2012), e34358.

[33] Agnik Saha, Mohammad Shahidul Kader, and Mohammad Masum. 2024. Unveiling User Engagement Patterns on Stack Exchange Through Network Analysis. *arXiv preprint arXiv:2409.08944* (2024).

[34] Betsy Sparrow, Jenny Liu, and Daniel M Wegner. 2011. Google effects on memory: Cognitive consequences of having information at our fingertips. *science* 333, 6043 (2011), 776–778.

[35] Stack Overflow. 2024. Badges - Stack Overflow Help Center. https://stackoverflow.com/help/badges Accessed: 2025-05-02.

[36] Stanford University Human-Centered Artificial Intelligence. 2025. *Artificial Intelligence Index Report 2025*. Technical Report. Stanford University, Stanford, CA.

[37] Susan Leigh Star and Anselm Strauss. 1999. Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer supported cooperative work (CSCW)* 8 (1999), 9–30.

[38] Sri Lakshmi Vadlamani and Olga Baysal. 2020. Studying software developer expertise and contributions in Stack Overflow and GitHub. In *2020 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 312–323.

[39] Molly McLure Wasko and Samer Faraj. 2005. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS quarterly* (2005), 35–57.

[40] Dennis M Wilkinson and Bernardo A Huberman. 2007. Cooperation and quality in wikipedia. In *Proceedings of the 2007 international symposium on Wikis*. 157–164.

[41] Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, and Yang Tang. 2023. A brief overview of ChatGPT: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica* 10, 5 (2023), 1122–1136.

[42] Lei Xu, Tingting Nian, and Luis Cabral. 2020. What makes geeks tick? A study of stack overflow careers. *Management Science* 66, 2 (2020), 587–604.