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 adapted or disengaged in response. In this study, we examine how contributor engagement shifts in these online knowledge communities, using Stack Overflow as our case site. We do this by clustering over 1.2 million 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 group with a matched control suggests heterogeneous trajectories. Some contributors, such as Steadfast Contributors and Lone Solvers, exhibited relatively stable participation or shifted toward writing longer posts. Others, including Peer-Conscious users, Solution Seekers, Viral Askers, and Thread Revivers, showed significantly steeper declines in the volume of questions and answers. These patterns may be consistent with a reconfiguration of participation rather than uniform disengagement. We discuss implications for the evolving dynamics of knowledge sharing as participation shifts in the context of generative AI.

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, 20 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 how different types of contributors have changed their engagement patterns following the emergence of ChatGPT: Who exhibits sustained activity, and who shows signs of disengagement over time? While prior research has documented a general decline in overall usage on Stack Overflow following the release of ChatGPT [9, 12, 13, 22, 31], it remains unclear how behavioral trends differ across user groups. Exploring these heterogeneous patterns may offer insights into the future of online knowledge sharing in the age of generative AI. Rather than focusing solely on aggregate usage trends, we examine individual-level behavioral variation by identifying distinct contributor profiles.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CSCW '26, USA
© 2025 ACM.
ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

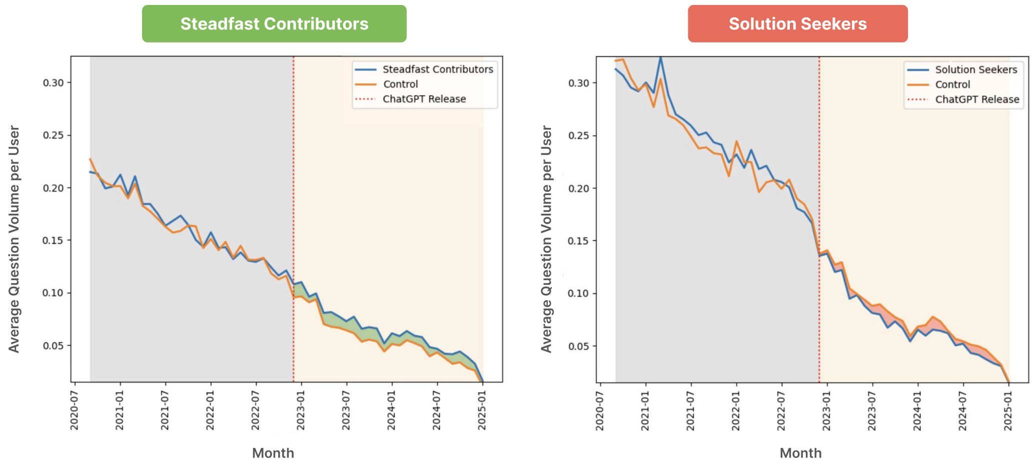

Fig. 1. Monthly question and answer volume per user for Steadfast Contributors and Solution Seekers, alongside their matched control groups. The vertical dashed line marks the public release of ChatGPT on November 30, 2022. Post-release divergence highlights distinct behavioral changes across user groups, reflecting heterogeneous patterns observed during the post-ChatGPT period.

To this end, we clustered over 1.2 million 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 [33]; 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 — 307,471 question askers and 296,235 answer contributors — and examined their engagement levels. For each user group, we constructed a matched control group by calculating propensity scores based on pre-treatment activity trends. We then conducted a difference-in-differences (DiD) analysis to compare 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 while overall engagement declined following the release of ChatGPT, user groups exhibited distinct trajectories. Certain contributors showed signs of relative resilience or adaptation, while others experienced significantly steeper declines in engagement than their matched controls. Steadfast Contributors—highly active users embedded in the platform's infrastructure—exhibited a smaller decline in question volume and a substantial increase in the length of their questions and answers. This pattern may be consistent with a shift in the way they contribute to the community, rather than strict disengagement. Similarly, Lone Solvers—users who answer questions even with limited recognition—did not differ significantly in volume trends from their matched controls, yet tended to write longer questions and answers on average.

In contrast, several contributor types showed significantly steeper declines in engagement relative to their matched controls. Peer-Conscious users—who tend to delete poorly received posts—exhibited a sharper drop in both question and answer volume. This may reflect discomfort with public contribution in an environment where AI tools can privately fulfill informational needs. Solution Seekers—users who proactively promote and incentivize their own questions—also experienced a significantly steeper decline in question volume. This group showed patterns consistent with reduced participation, which could align with the availability of tools like ChatGPT, as their participation appears to be driven primarily by the instrumental goal of obtaining solutions, rather than by intrinsic motivation or community identification. Thread Revivers—users with a history of answering long-dormant questions—showed sharp reductions in both question and answer volume as well.

These behavioral shifts carry several implications. First, our findings extend prior work by suggesting that while overall content volume has declined following the release of ChatGPT, the average length of posts has increased. This pattern aligns with earlier studies [9, 12, 13, 22] and may reflect a shift toward more complex contributions as simpler questions are increasingly handled by AI tools. Second, by identifying canonical user profiles and analyzing their differential trajectories, our study reveals a reconfiguration of user engagement rather than a uniform disengagement. Contributors with intrinsic or identity-based motivations were associated with relatively stable activity, while those with instrumental or socially sensitive motivations were associated with steeper declines in activity. Finally, the reduced engagement of contributors whose work, while often low-visibility, is essential to sustaining archival completeness, such as Thread Revivers, coincides with what could be emerging vulnerabilities in the community's ability to maintain the breadth and depth of its collective knowledge base. Taken together, this study sheds light on how generative AI is reshaping the online knowledge community, highlighting which users are changing, how their 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 which types of contributors exhibit milder versus sharper declines in engagement levels in response to generative AI, revealing heterogeneous behavioral adaptation across user groups.
- We offer perspectives on how the meaning and practice of online knowledge sharing may be evolving in the era of generative AI.
- Methodologically, we present an application of topic modeling for clustering users to derive canonical user prototypes from 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 [36]. 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 [25, 28, 33, 40]. Expertise is organized largely around tags, with users gravitating toward specific technical domains (e.g., Python, JavaScript, Android). These domain-focused communities are notably stable over time, with little evidence of cross-domain migration [27]. Specialization in a narrow range of tags is positively associated with both answer quality and long-term reputation growth [24].

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 [38]. These tasks often constitute forms of "invisible work" that are critical to the integrity, reliability, and long-term accessibility 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, 22, 31]. 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 [31]. 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. [?] 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 [34]. 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. [23] 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 [33]. 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 [39], 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 included 307,471 users with valid question activity—hereafter referred to as question askers. Similarly, 296,235 users with valid answer activity were retained—hereafter referred to as answer contributors. These datasets form the foundation for our clustering and subsequent analyses.

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.

3.2 User Clustering

To identify meaningful user groups based on their behavioral patterns, we clustered the 1,277,423 users who had earned at least one of the selected 31 badges prior to the release of ChatGPT.

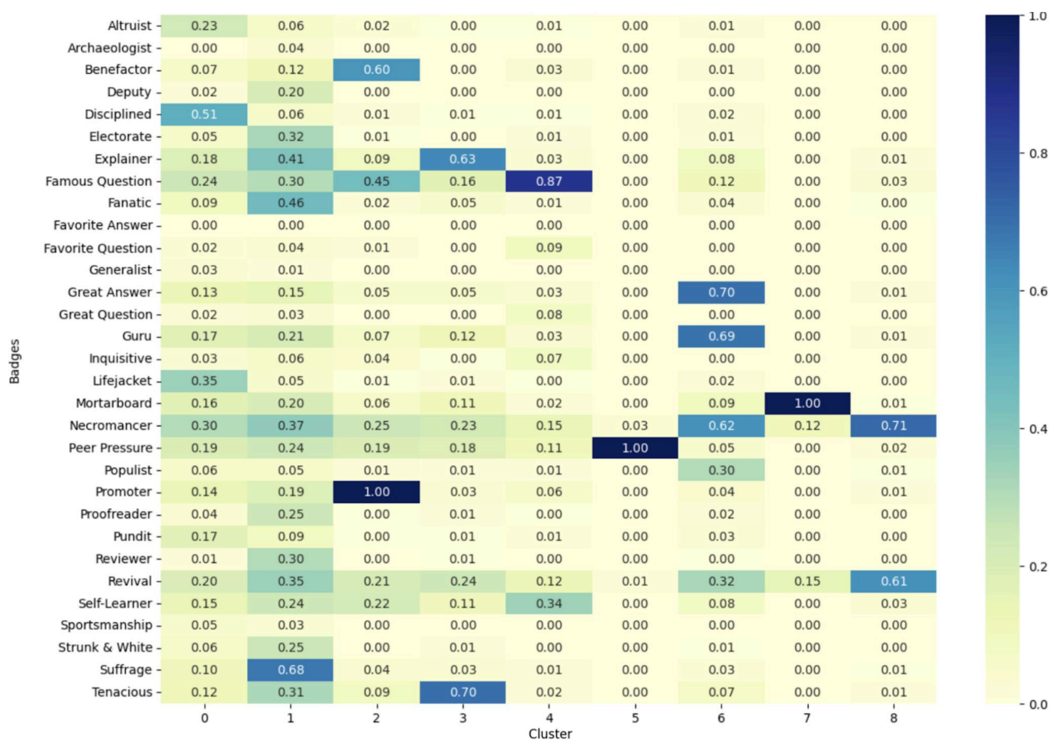

Fig. 2. Final badge clustering results using our two-step method with guided LDA. Each row represents a badge, and each column represents a user cluster. Color intensity reflects the proportion of users in a cluster who earned the badge (1.0 = all users, 0.0 = none).

We adopted a topic modeling approach using Latent 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 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 1,277,423 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, 21], 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 0.35.

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	Transient Contributors	These users participated briefly but retracted their posts, even when the posts were well-received (Disciplined). This pattern points to shallow engagement, possibly driven by a lack of interest in long-term contribution to the community.	Disciplined (51)	10,739
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 (68), Fanatic (46), Explainer (41), Necromancer (37)	67,129

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 (45)	50,722
3	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 (70), Explainer (63)	37,220
4	Viral Askers	These users have posted one or more questions that attracted widespread attention (Famous Question).	Famous Question (87), Self-Learner (34)	447,793
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)	232,492
6	Expert Answerers	These high-reputation users consistently provide top-quality answers (Great Answer, Guru) and specialize in reviving old questions with valuable insights (Necromancer). Their contributions often stand out in technical excellence and community impact.	Great Answer (70), Guru (69), Necromancer (62)	32,284
7	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)	3,478
8	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 (71), Revival (61)	395,566

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 control group of users.

3.3.1 Matched Control Construction. We constructed a matched control group for each treated cluster using a one-to-one propensity-score matching. The goal of this procedure is to ensure that the control group closely resembles the treated group (i.e., the user cluster) in terms of pre-treatment behavioral trends, enabling a valid difference-in-differences (DiD) comparison [6]. For each of the nine clusters, we applied the following steps.
Step 1: Identifying treated users. We defined the treated group as the set of users belonging to a specific cluster. The control group was sampled from the remaining users who did not belong to that cluster.
Step 2: Constructing per-user features. For all users, we computed a set of summary statistics based on their activity during the pre-treatment 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 for each user. These seven variables constituted the matching covariates.
Step 3: Estimating propensity scores. We estimated propensity scores using logistic regression, predicting the likelihood of a user belonging to the treated group based on their standardized features. We then applied a common support condition by restricting the donor pool (potential controls) to users whose propensity scores fell within a narrow caliper (±0.05) of the treated group's score range.
Step 4: One-to-one matching without replacement. Within the reduced donor pool, we paired each treated user with exactly one control user. By default, we used the Mahalanobis distance computed over the covariates. This matching minimizes the total multivariate distance between pairs while accounting for covariate covariance. For only one cluster ($K = 4$) with more than 119,000 treated users, a full distance matrix was pointless, so we instead used greedy nearest-neighbor matching on the propensity score. The result is a matched sample with equal numbers of treated and control 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 control 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 samples. Table 3 shows the standardized difference of covariates before and after matching. 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].

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

$$Y_{it} = \beta_0 + \beta_1 \cdot \text{Treated}_i + \beta_2 \cdot \text{Post}_t + \beta_3 \cdot (\text{Treated}_i \times \text{Post}_t) + \epsilon_{it} \tag{1}$$

Table 3. Standardized differences in the number of questions and answers per cluster, before and after matching.

		Question Askers		Answer Contributors	
K	Cluster	Before	After	Before	After
0	Transient Contributors	0.025	0.004	0.220	0.004
1	Steadfast Contributors	0.048	0.004	0.345	0.083
2	Solution Seekers	0.262	0.009	-0.043	0.008
3	Lone Solvers	-0.122	0.004	0.439	0.008
4	Viral Askers	0.140	0.058	-0.180	0.007
5	Peer-Conscious	-0.132	-0.001	-0.182	0.008
6	Expert Answerers	-0.336	0.000	-0.105	0.007
7	Peak Performers	-0.073	0.002	0.093	0.015
8	Thread Revivers	-0.312	0.003	-0.206	0.009

Here, Y_{it} is the outcome variable for user i at month t, $\texttt{Treated}_i$ is a binary indicator for whether the user belongs to the treated group (i.e., a user cluster), $\texttt{Post}_t$ is a binary indicator for whether the observation occurs after the public release of ChatGPT (November 30, 2022), and the interaction term $\texttt{Treated} \times \texttt{Post}$ captures the DiD effect of interest. The coefficient β_3 represents the treatment effect: the differential change in the outcome variable for the treated group compared to the control group following the introduction of ChatGPT. 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 ordinary least squares (OLS), with standard errors clustered at the user level to account for repeated observations of the same individuals over time. This clustering adjustment helps correct for potential heteroskedasticity and autocorrelation within users' behavioral trajectories.

We fit each of the two models separately for each user cluster and its matched control group. For each model, we report the estimated coefficient β_3, its statistical significance, and confidence intervals to assess whether the post-treatment behavior of each cluster differs meaningfully from its counterfactual.

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

3.4 Parallel Trend Assumption

A key identification requirement for DiD is the *parallel trends assumption*, which posits that, in the absence of treatment, the outcome trajectories of different groups would have followed similar paths [6].

To evaluate this assumption, we estimated the following linear model restricted to the pre-treatment period.

$$Y_{it} = \gamma_0 + \gamma_1 \cdot \text{Treated}_i + \gamma_2 \cdot \text{Month}_t + \gamma_3 \cdot (\text{Treated}_i \times \text{Month}_t) + \epsilon_{it} \tag{2}$$

In this specification, Y_{it} denotes user i's average monthly questions or answers at time t, $Treated_i$ indicates whether the user i belongs to the treatment group, and $Month_t$ is an indicator of month within the pre-treatment window. Here, γ_1 captures any initial baseline differences, γ_2 reflects the pre-treatment trend for the control group, and γ_3 represents the slope difference between treated and controls in the pre-treatment period.

Table 4 reports the estimated γ_3 coefficients for each cluster. For both questions and answers, eight out of nine groups exhibit γ_3 values that are statistically indistinguishable from zero ($p > 0.05$), indicating no detectable divergence in pre-treatment slopes and thus supporting the parallel trend assumption.

Among question askers, the sole exception is the Viral Askers ($K = 4$), for whom the pre-treatment slope difference is statistically significant ($p < 0.001$), though the magnitude is minimal $\gamma_3 = -0.0006$. Given the large

Table 4. Difference in pre-treatment slopes between treated and control users.

		Question Askers			Answer Contributors		
K	Cluster	# Users	γ_3	p	# Users	γ_3	p
0	Transient Contributors	3,083	-4.594×10^{-5}	0.918	3,030	-0.0002	0.939
1	Steadfast Contributors	24,502	4.809×10^{-6}	0.976	27,890	-0.0029^{**}	0.006
2	Solution Seekers	26,320	-5.045×10^{-5}	0.803	17,675	2.266×10^{-5}	0.958
3	Lone Solvers	10,786	-1.121×10^{-5}	0.944	15,693	0.0002	0.854
4	Viral Askers	119,238	-0.0006^{***}	< 0.001	78,377	4.617×10^{-5}	0.658
5	Peer-Conscious	61,002	0.0001	0.175	29,290	0.0002	0.239
6	Expert Answerers	4,084	3.199×10^{-5}	0.846	8,122	-8.772×10^{-5}	0.825
7	Peak Performers	751	3.658×10^{-5}	0.959	1,075	0.0002	0.931
8	Thread Revivers	57,705	-2.846×10^{-6}	0.954	114,202	-4.576×10^{-6}	0.952

Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

sample size ($N = 119{,}238$), this significance likely reflects high statistical power rather than a substantively meaningful divergence. We therefore consider the practical deviation negligible and retain this group in the DiD analysis, while interpreting their results with appropriate caution.

Among answer contributors, Steadfast Contributors ($K = 1$) show a statistically significant pre-treatment slope difference ($p < 0.01$), again with a small magnitude $\gamma_3 = -0.0029$. We attempted to improve match quality by tightening calipers to zero but were unable to achieve adequate balance, reflecting this group's extremeness in baseline activity. Accordingly, we flag this case as a violation of the parallel trends assumption and interpret subsequent treatment effects conservatively.

4 Results

4.1 Question Volume

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

4.1.1 Universal Decline. The β_2 estimate is negative in every cluster, indicating that all control groups reduced their question-asking after ChatGPT's release. Because every interaction term β_3 is small relative to β_2, all treated groups also declined. These results are consistent with prior work [9, 13, 22].

4.1.2 Cluster-specific Divergence. Steadfast Contributors experienced a significantly smaller drop than their matched controls. With an extra +0.008 questions per user-month ($p < 0.01$), they offset about 7% of the control decline (−0.109).

By contrast, Solution Seekers ($\beta_3 = -0.010,\ p < 0.01$), Peer-Conscious users ($\beta_3 = -0.003,\ p < 0.01$), and Thread Revivers ($\beta_3 = -0.004,\ p < 0.001$) experienced disproportionately larger contractions in question volume, approximately 4.2%, 2.5% and 5.8% greater declines, respectively, relative to their matched controls. In concrete terms, these groups shed an additional 2-6% of activity on top of the overall downward trend.

Viral Askers ($\beta_3 = -0.003,\ p < 0.05$) likewise exhibited a steeper drop, about 2.1% more than controls. However, given their imperfect pre-treatment matching, we cannot conclusively attribute this difference to the post-ChatGPT period rather than to lingering baseline disparities.

For the remaining clusters, the interaction confidence intervals span zero, and the effects are statistically indistinguishable from null ($p > 0.05$), implying that their post-ChatGPT contraction is statistically and practically indistinguishable from their matched controls.

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

Cluster	β_2		β_3			Effect size (%)
	Estimate	95% CI	Estimate	95% CI	p	
Transient Contributors	-0.102^{***}	$[-0.114, -0.090]$	-0.005	$[-0.021, 0.012]$	0.573	-3.1
Steadfast Contributors	-0.109^{***}	$[-0.114, -0.105]$	0.008^{*}	$[0.002, 0.013]$	0.012	4.6
Solution Seekers	-0.169^{***}	$[-0.174, -0.164]$	-0.010^{**}	$[-0.017, -0.003]$	0.005	-4.2
Lone Solvers	-0.070^{***}	$[-0.073, -0.067]$	-0.002	$[-0.006, 0.003]$	0.482	-1.4
Viral Askers	-0.110^{***}	$[-0.112, -0.109]$	-0.003^{*}	$[-0.006, 0.001]$	0.013	-2.1
Peer-Conscious	-0.088^{***}	$[-0.090, -0.087]$	-0.003^{**}	$[-0.005, -0.001]$	0.003	-2.5
Expert Answerers	-0.029^{***}	$[-0.033, -0.026]$	-0.001	$[-0.005, 0.003]$	0.647	-1.6
Peak Performers	-0.088^{***}	$[-0.101, -0.074]$	-0.007	$[-0.026, 0.013]$	0.496	-5.4
Thread Revivers	-0.041^{***}	$[-0.042, -0.040]$	-0.004^{***}	$[-0.006 - 0.003]$	< 0.001	-5.8

Notes: β_2 captures the pre/post change for the matched control users; β_3 is the DiD estimate for the treated cluster. Effect size = β_3/β_0, expressed as a percentage of the pre-ChatGPT control baseline (β_0). Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

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, all treated clusters exhibited significant increases in average question length post-ChatGPT, with gains ranging from +51 characters for Transient Contributors to +377 characters for Thread Revivers.

Among control groups, eight clusters also showed increases in question length, ranging from +20 to +274 characters, five of which were statistically significant. However, the control group for Thread Revivers diverged from this trend, showing a substantial and statistically significant drop of 151 characters ($p < 0.001$).

4.2.2 Cluster-specific Divergence. Several clusters experienced a significantly steeper increase in question length compared to their matched controls after ChatGPT's release. Peer-Conscious users showed increase with an extra +218.17 characters ($p < 0.001$). Steadfast Contributors had extra +99 characters ($p < 0.001$), Lone Solvers had extra +91 characters ($p < 0.05$), and Solution Seekers had extra +59 characters ($p < 0.05$), representing 5.1%, 4.6%, and 2.9% steeper increase relative to their baselines.

On the other hand, Viral Askers had an increase by 91 characters relative to controls ($p < 0.001$).

Other clusters showed no reliable difference in their adaptation ($p > 0.05$).

4.3 Answer Volume

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

4.3.1 Universal Decline. Every control group saw a highly significant drop in answers per month (all $\beta_2 < 0, p < 0.001$). Likewise, all treated groups also show a decline —from −0.074 answers per month for Viral Askers to −0.723 for Lone Solvers.

4.3.2 Cluster-specific Divergence. Steadfast Contributors experienced a steeper decline in answer volume than their matched controls, with an additional decrease of −0.097 answers per user-month ($p < 0.001$). However, this estimate may be partially confounded by baseline differences due to the violation of the parallel trend assumption, rather than reflecting purely post-ChatGPT effects. Furthermore, Steadfast Contributors began from an exceptionally high baseline: before ChatGPT, they averaged 1.030 answers per user-month, compared to the overall pre-ChatGPT mean of 0.275 ($SD = 2.330$). After ChatGPT, they averaged 0.343, compared to the overall post-ChatGPT mean of 0.088 ($SD = 1.127$), remaining well above average.

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

Cluster	β_2 Estimate	β_2 95% CI	β_3 Estimate	β_3 95% CI	p	Effect size (%)
Transient Contributors	129.93*	[17.25, 242.61]	51.27	[−118.28, 220.82]	0.553	+2.7
Steadfast Contributors	121.83***	[79.78, 163.88]	99.16***	[39.63, 158.69]	< 0.001	+5.1
Solution Seekers	34.65	[−2.95, 72.25]	59.11*	[4.25, 113.97]	0.035	+2.9
Lone Solvers	19.72	[−42.02, 81.46]	90.67*	[1.20, 180.13]	0.047	+4.6
Viral Askers	274.32***	[253.21, 295.42]	−90.69***	[−118.78, −62.59]	< 0.001	−4.8
Peer-Conscious	220.45***	[193.48, 247.42]	218.17***	[172.06, 264.29]	< 0.001	+12.5
Expert Answerers	154.23***	[61.78, 246.69]	20.47	[−115.83, 156.78]	0.768	+1.2
Peak Performers	105.96	[−89.92, 301.84]	34.73	[−250.08, 319.54]	0.811	+1.9
Thread Revivers	189.46***	[159.62, 219.29]	36.49	[−6.45, 79.42]	0.096	+2.0

Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

Beyond Steadfast Contributors, we also find that Viral Askers ($\beta_3 = -0.008,\ p < 0.001$), Peer-Conscious users ($\beta_3 = -0.010,\ p < 0.001$), and Thread Revivers ($\beta_3 = -0.010,\ p < 0.001$) contracted more sharply than their matched controls, each shedding an additional 5–9% of activity relative to the general downward trend. For the remaining clusters, the effects are statistically indistinguishable from null ($p > 0.05$), implying that their post-ChatGPT contraction is indistinguishable from their matched controls.

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

Cluster	β_2 Estimate	β_2 95% CI	β_3 Estimate	β_3 95% CI	p	Effect size (%)
Transient Contributors	−0.462***	[−0.527, −0.397]	0.061	[−0.041, 0.164]	0.242	+8.6
Steadfast Contributors	−0.575***	[−0.597, −0.554]	−0.097***	[−0.138, −0.055]	< 0.001	−12.1
Solution Seekers	−0.153***	[−0.165, −0.142]	−0.014	[−0.029, 0.001]	0.060	−6.4
Lone Solvers	−0.723***	[−0.755, −0.691]	−0.036	[−0.082, 0.009]	0.115	−3.7
Viral Askers	−0.074***	[−0.078, −0.071]	−0.008***	[−0.012, −0.004]	< 0.001	−6.2
Peer-Conscious	−0.075***	[−0.078, −0.071]	−0.010***	[−0.015, −0.006]	< 0.001	−8.8
Expert Answerers	−0.106***	[−0.116, −0.096]	0.009	[−0.005, 0.023]	0.222	+5.3
Peak Performers	−0.268***	[−0.315, −0.220]	−0.021	[−0.083, 0.040]	0.498	−6.0
Thread Revivers	−0.076***	[−0.079, −0.074]	−0.007***	[−0.010, −0.004]	< 0.001	−5.6

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. Most treated clusters exhibited significant increases in average answer length post-ChatGPT, with gains ranging from +37 characters for Lone Solvers to +147 characters for Peer-Conscious users. Specifically, Transient Contributors (+60 chars, $p = 0.041$), Solution Seekers (+41 chars, $p < 0.001$), Lone Solvers (+37 chars, $p < 0.001$), Viral Askers (+73 chars, $p < 0.001$), Peer-Conscious (+147 chars, $p < 0.001$), and Thread Revivers (+75 chars, $p < 0.001$) all showed significant upward shifts. Expert Answerers and Peak Performers did not experience significant changes in the length of the answer ($p > 0.05$).

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

Cluster	β_2		β_3			Effect size (%)
	Estimate	95% CI	Estimate	95% CI	p	
Transient Contributors	59.61*	[2.47, 116.75]	40.70	[−39.28, 120.69]	0.319	+4.4
Steadfast Contributors	36.95***	[21.07, 52.82]	53.46***	[31.46, 75.46]	<0.001	+5.5
Solution Seekers	40.99***	[18.10, 63.88]	30.76	[−2.32, 63.84]	0.068	+3.6
Lone Solvers	37.45***	[18.20, 56.70]	39.00**	[11.09, 66.91]	0.006	+4.0
Viral Askers	72.81***	[61.72, 83.90]	−15.32	[−30.86, 0.23]	0.053	−1.9
Peer-Conscious	147.05***	[129.81, 164.29]	−50.54***	[−76.74, −24.34]	<0.001	−7.3
Expert Answerers	17.46	[−15.19, 50.12]	32.11	[−12.83, 77.06]	0.161	+3.7
Peak Performers	−11.04	[−122.06, 99.99]	−32.29	[−179.62, 115.04]	0.667	−3.1
Thread Revivers	75.18***	[66.04, 84.32]	−19.45**	[−32.46, −6.45]	0.003	−2.4

Significance: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

4.4.2 Cluster-specific Divergence. Several clusters diverged significantly from their matched controls in how answer length changed after ChatGPT's release. Peer-Conscious users did experience an absolute increase in answer length, but their gain was significantly smaller than that of their matched controls, amounting to -50 characters in relative terms ($p < 0.001$, −7.3%). Similarly, Thread Revivers also increased their answer length overall, but by a smaller margin than their controls, corresponding to −19 characters in relative terms ($p = 0.003$, −2.4%). In contrast, Lone Solvers exhibited a significantly steeper rise relative to their controls, contributing answers that were on average +39 characters longer ($p = 0.006$, +4.0%).

Note that these DiD coefficients represent additional changes relative to matched controls and should therefore be interpreted as incremental to the platform-wide decline, not as absolute levels. Aggregating to the cluster level clarifies the practical scale. For example, for Peer-Conscious users, a reduction of 50 characters per user-month translates into more than 1.46 million fewer characters in total relative to matched controls (29,290 users × -50 characters). This is incremental to the post-period platform decline.

Steadfast Contributors also demonstrated a steeper increase in answer length, with an average gain of +53 characters compared to their matched controls ($p < 0.001$, +5.5%). However, this estimate should be interpreted with caution, as the Zealot group violated the parallel trends assumption in the pre-treatment period. As such, the observed difference may be attributable to pre-existing baseline disparities rather than changes that emerged exclusively after ChatGPT's release.

Other clusters did not differ significantly from their controls in post-ChatGPT answer length adaptation ($p > 0.05$).

5 Discussion

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 relative stability in contribution behavior despite an external technological shift. Steadfast Contributors—highly active contributors embedded in the norms and infrastructure of the platform—and Lone Solvers—contributors who are likely to be less sensitive to community feedback or recognition—showed relatively stable engagement compared to their matched controls.

Steadfast Contributors offset roughly 7% of the control group's decline in question volume and boosted both question (+99 chars) and answer length (+53 chars). This pattern suggests not strict disengagement, but role refinement: these users may be adapting the way they participate. 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 [29]. In the case of Steadfast Contributors, motivation may stem not only from obtaining information

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

Cluster	Description	Question Volume	Question Length	Answer Volume	Answer Length
Transient Contributors	Briefly active users who deleted posts despite positive feedback.	-3.1 %	+2.7 %	+8.6 %	+4.4 %
Steadfast Contributors	Consistently active and deeply engaged across various activities.	**+4.6 %**	**+5.1 %**	**-12.1 %**	**+5.5 %**
Solution Seekers	Promoted and enhanced their questions to get better answers.	**-4.2 %**	+2.9%	-6.4 %	+3.6 %
Lone Solvers	Persistently answered questions without needing recognition.	-1.4 %	+4.6 %	-3.7 %	**+4.0 %**
Viral Askers	Their questions gained widespread attention.	**-2.1 %**	**-4.8 %**	**-6.2 %**	-1.9 %
Peer-Conscious	Remove poorly received posts to align with community norms.	**-2.5 %**	**+12.5 %**	**-8.8 %**	**-7.3 %**
Expert Answerers	Provided consistently top-quality, technical answers.	-1.6 %	+1.2 %	+5.3 %	+3.7 %
Peak Performers	Contributed high-impact content in intense bursts.	-5.4 %	+1.9 %	-6.0 %	-3.1 %
Thread Revivers	Re-engaged with and updated old, dormant threads.	**-5.8 %**	+2.0 %	**-5.6 %**	**-2.4 %**

Notes: Statistically significant results are indicated in **bold**. Green highlights represent a relative increase, while red indicates a relative decrease compared to the matched controls.

but from identity-based and intrinsic drivers, such as contributing to a shared resource or reinforcing one's social role [37]. It is possible that their participation is sustained not in spite of generative AI, but because their connection to the community serves broader personal or social functions that AI tools do not replace.
Lone Solvers exhibited no significant difference in volume decline relative to their matched controls, yet their posts grew longer on average (+91 characters on questions; +39 on answers). In other words, even though they are posting less frequently, they tend to write more when they do contribute. This shift suggests a form of adaptation—not in how often they engage, but in the way they engage. Lone Solvers may be responding to the changing platform environment, shaped in part by the rise of generative AI, by posting longer questions and answers. Their tendency to write longer content aligns with a focus on problem-solving for its own sake rather than external feedback. In this sense, their post-ChatGPT behavior may reflect a sustained commitment to offering solutions, even as broader participation patterns shift.

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 control 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 question volume—more severe than that of their matched control 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 [32], 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 control 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 [30, 35]. 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.

Viral Askers are users with a history of asking questions that attracted widespread attention. Following the release of ChatGPT, this group experienced a significantly steeper drop in both questions asked and answers given—exceeding the decline seen in their matched controls. This divergence may indicate that AI's instant, high-quality responses to broadly popular queries undercut the very motivation that drove these users to engage. Once the most sought-after information could be resolved seamlessly by ChatGPT, the incentive to solicit and share solutions on the platform may have diminished.

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, 22]. While the overall volume of questions and answers declined, the average length of posts increased. One possible explanation is that simple questions are now easily answered by AI, leaving longer and possibly more complex questions to be posted on the platform. This interpretation aligns with prior findings suggesting that knowledge communities may retain an advantage in addressing nuanced and context-heavy problems [9, 23].

Second, we contribute a methodological approach to this domain by identifying canonical user prototypes through unsupervised clustering based on behavioral data. Our analysis reveals that the impact of generative AI is not uniform across different contributors: while groups such as Peer-Conscious users, Solution Seekers, Thread Revivers, and Viral Askers show steeper declines in activity, others like Steadfast Contributors and Lone Solvers exhibit comparatively stable or adaptive participation. This segmentation helps reveal how participation patterns may be shifting, rather than uniformly declining, offering insights into the evolving landscape of knowledge sharing.

Lastly, the sharp 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-treatment 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 controls are badge-earning users, our findings reflect behavioral patterns within the subpopulation of actual contributors rather than the average Stack Overflow user.

Furthermore, 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 post-treatment engagement patterns through a difference-in-differences analysis, we find that responses to generative AI vary across user roles. While overall activity declined, certain groups, such as Steadfast Contributors and Lone Solvers, exhibited signs of adaptation through either a milder decline in question volume or increased content length. In contrast, contributors such as Peer-Conscious users, Solution Seekers, Viral Askers, and Thread Revivers showed significantly steeper declines in engagement compared to their matched controls.

These heterogeneous responses suggest that generative AI is not simply reducing participation, but rather may be altering patterns of participation across roles. Contributors with intrinsic or identity-based motivations may be more resilient, whereas users driven by instrumental goals or socially sensitive dynamics may be more susceptible to disengagement. The decline of structurally important roles may have long-term implications for the completeness 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. Rather than assessing only aggregate changes, this study aims to understand who is adapting, how they are adapting, 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] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

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

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

[29] 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.

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

[31] 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).

[32] 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.

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

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

[35] 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.

[36] 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.

[37] 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.

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

[39] 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.

[40] 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.