Examining Changes in Good First Issue Practices and Newcomer Pull Request Characteristics in Popular OSS Projects

Anonymous Author(s)

ABSTRACT

Open-source software (OSS) projects rely on effective newcomer onboarding to sustain their communities. Many projects use "good first issue" (GFI) labels to highlight beginner-friendly tasks. As development practices continue to evolve, understanding how these onboarding mechanisms change over time is important for both maintainers and researchers. This study analyzes 406,826 issues and 1,117 PRs addressing GFIs across 37 popular GitHub repositories (30 of which use GFI labels) over a four-year period from July 2021 to June 2025. We find that while the proportion of issues with GFI labels remained stable during the first three years, it underwent a structural decline beginning in January 2024, with substantial variation across projects not explained by repository age or programming language. Despite this supply-side decline, newcomer engagement with GFI issues remains stable at approximately 27%, suggesting that GFI labels maintain consistent attractiveness. Examining the outcomes of this engagement, we find that the merge rate of newcomer GFI PRs declined from 61.9% to 42.2%. Initial PR characteristics such as description length and code size show no significant association with merge outcomes, indicating that success is not predicted by the quantitative characteristics of the initial submission alone. Together, these findings indicate that both the supply and success of GFI-based onboarding are declining in parallel, likely reflecting reduced maintainer investment in newcomer support.

KEYWORDS

Open source software, Newcomer onboarding, Good first issue, GitHub

1 INTRODUCTION

The sustained development of open-source software (OSS) projects depends critically on the influx and retention of new contributors [1, 2]. However, newcomers often face significant technical and social barriers when attempting to make their first contribution, including difficulty finding suitable tasks, understanding complex codebases, and navigating unfamiliar development processes [1, 3].

To address these challenges, many OSS projects have adopted the practice of labeling certain issues as "good first issue" (GFI) to identify tasks suitable for beginners. Prior research has examined the usage patterns of GFI labels and proposed automated recommendation methods [4, 5]. However, Tan et al. [6] showed that many GFIs are not resolved by newcomers, indicating that challenges remain regarding the effectiveness of GFI labeling practices.

In recent years, the software development landscape has undergone substantial changes, including the widespread adoption of remote work and the emergence of generative AI tools and Large Language Models (LLMs) that are reshaping how developers write and review code. Understanding how GFI practices and newcomer behavior have evolved over time is important for maintaining healthy OSS communities.

However, **longitudinal trends in GFI practices and the behavior of newcomers who engage with GFI issues remain unclear**. While prior work has examined GFI effectiveness and recommendation methods at specific points in time, our understanding of how GFI practices have evolved over recent years—and of how the characteristics and success rates of GFI-related contributions have changed—remains limited.

In this study, we analyze four years (July 2021 to June 2025) of GFI practices across 37 popular OSS projects on GitHub. We address the following research questions:

RQ1: How have GFI practices and newcomer engagement changed over the four-year period?

We examine the trends in GFI ratio and newcomer engagement rates, investigating how these patterns have evolved over time.

Understanding both the availability of GFI-labeled issues and the outcomes when newcomers engage with them is essential for a comprehensive picture of the GFI ecosystem.

RQ2: How do task-type labels of GFI issues relate to newcomer PR merge outcomes, and what factors are associated with merge success?

Given the GFI trends identified in RQ1, we further investigate what happens when newcomers engage with GFIs. We classify GFI issues into task types (Bug, Feature, Documentation, Other) based on their labels and analyze how merge rates differ by task type and over time. We also examine PR-level factors associated with merge success.

Through an analysis of 406,826 issues (identifying 3,300 GFI-labeled issues) and 1,117 GFI PRs, we find that the GFI ratio underwent a structural decline beginning in January 2024 (Pettitt test, $p < 0.001$), with substantial cross-project variation. Despite this decline in GFI availability, newcomer engagement remains stable at approximately 27%. However, the merge rate of newcomer GFI PRs also declined, suggesting that the challenges extend beyond supply to the success of onboarding interactions.

2 RELATED WORK

Newcomer Onboarding in OSS. Steinmacher et al. [1, 3] systematically classified the barriers that newcomers face when joining OSS projects, highlighting the importance of social as well as technical barriers. Subramanian et al. [7] analyzed the characteristics of newcomers' first contributions, revealing that approximately half of them address bug fixes.

Turzo et al. [8] evaluated the effectiveness of onboarding recommendations and showed that effective strategies vary by project. Steinmacher et al. [9] proposed FLOSScoach, a portal for newcomers, and demonstrated its effectiveness in reducing orientation barriers.

Good First Issues and Task Recommendation. Tan et al. [6] were the first to systematically analyze the use of the Good First Issue (GFI) label on GitHub, showing that many GFIs are not resolved by newcomers. Xiao et al. [4] and Huang et al. [5] proposed methods for automatically recommending GFIs using machine learning. Santos et al. [10] revealed a mismatch in task selection strategies between newcomers and existing developers. Xiao et al. [11] proposed a personalized task recommendation method based on contributor background.

Mentoring and Community Support. Steinmacher et al. [12] and Balali et al. [13] identified the challenges and strategies of task recommendation from the perspective of OSS mentors. Cao et al. [14] showed that simply labeling GFIs is insufficient and that direct support from mentors is crucial for newcomer success. Guizani et al. [15] proposed a maintainer dashboard to support the attraction of new contributors. Setiawan et al. [**?**] analyzed how initial PR characteristics relate to newcomer retention, providing insight into the factors that influence early contribution success.

Building on this prior work, our study empirically analyzes longitudinal trends in GFI usage and newcomer engagement over a four-year period. While prior studies examined GFI effectiveness at specific points in time, our work contributes by analyzing temporal trends using time-series methods.

3 METHODOLOGY

Figure 1 illustrates the overview of our research methodology. We selected our target projects from the top 50 most-starred repositories on GitHub, collected issues and newcomer pull requests, and analyzed GFI label trends and newcomer engagement (RQ1) and PR characteristics with merge success factors (RQ2).

3.1 Repository Selection

To focus on projects with active contributions, we only selected repositories with 50 or more pull requests per year. Furthermore, we excluded non-software projects (e.g., tutorial collections, learning resources) and projects that have disabled GitHub Issues (e.g., django/django), resulting in a final set of 37 software repositories for analysis, of which 30 used GFI labels.

The selected repositories are diverse, with the distribution of primary programming languages being TypeScript (9 projects, 24.3%), C++ (5, 13.5%), JavaScript (5, 13.5%), Python (5, 13.5%), Rust (3, 8.1%), and others (10, 27.0%).

3.2 Data Collection

3.2.1 Issue Data Collection (RQ1). To answer RQ1, we used the GitHub GraphQL API to collect all GFI-labeled issues from July 2021 to June 2025. To identify GFIs, we followed the method adopted by Turzo et al. [8], combining the label list presented by Tan et al. [6] with newcomer contribution guidelines [16]. For each issue, we recorded its creation date, label information, and closed state.

To calculate the GFI ratio—defined as the proportion of GFI-labeled issues among all issues created in a given month—we collected 406,826 total issues and identified 3,300 issues with GFI labels. Additionally, to analyze newcomer engagement with GFI tasks, we collected pull requests addressing these GFI issues.

3.2.2 Pull Request Data Collection (RQ2). To answer RQ2, we collected pull requests that address issues with GFIs (hereafter referred to as GFI PRs) using the GitHub GraphQL API. We define a *newcomer* as a contributor submitting their *first-ever pull request to a specific repository*—not their first contribution to OSS in general. Newcomers were therefore identified on a per-repository basis: we extracted the first-ever pull request submitted by each user to a given repository between July 2021 and June 2025. For each pull request, we retrieved the PR number, title, body, creation date, merge date, state (MERGED, CLOSED, OPEN), lines added, lines deleted, number of changed files, commit count, review comment count, and label information. For the description, we measured the length of substantive user-written content after removing HTML comments from PR templates. To ensure data quality, we filtered out bots and deleted accounts using the author_type field from the GitHub API, retaining only those where author_type was 'User'. All data was collected via the GitHub API in November 2025, approximately five months after the end of the study period. This observation buffer exceeds the 95th percentile of time-to-merge among merged PRs (approximately 80 days), ensuring that PRs created near the end of the analysis window had sufficient time to be reviewed and resolved. For merge rate calculations, we treated all unmerged PRs (including 68 still-open PRs) as unmerged. For insertions and deletions, which exhibited highly skewed distributions, we applied a log transformation.

For time-series comparison, we divided the four-year period into 12-month analysis years: Y1 (Jul 2021–Jun 2022), Y2 (Jul 2022–Jun 2023), Y3 (Jul 2023–Jun 2024), and Y4 (Jul 2024–Jun 2025). We classified each PR into a task type based on the labels of its referenced GFI issue: Bug (label contains "bug"), Feature ("feature" or "enhancement"), Documentation ("doc"), and Other (none of the above). Word-boundary matching for "bug" prevents false positives from area labels such as "debug." Of the 1,117 PRs, 1,070 (95.8%) matched exactly one task type and were classified automatically. The remaining 47 PRs (4.2%) matched multiple types; the first author manually classified each case by distinguishing type labels (e.g., "type: bug", "C-enhancement", "documentation") from area/module labels (e.g., "addon: docs", "A-docs", "module: docs"). For example, a PR with "type: bug" and "docs" (area) was classified as Bug, while a PR with both "enhancement" and "documentation" as type labels figures/method2.pdf

Figure 1: Overview of the study method.

was classified as Documentation. The full list of 47 overlapping cases with their manual classifications is included in the replication package.

The final dataset consists of 43,906 first pull requests from newcomers and 1,117 GFI PRs.

To control for multiple comparisons, we applied Holm-Bonferroni correction (controlling the family-wise error rate) within each analysis table, and Benjamini-Hochberg correction (controlling the false discovery rate at $\alpha = 0.05$) for the 30 individual repository-level trend tests.

4 RESULTS

4.1 RQ1: How have GFI practices and newcomer engagement changed over time?

4.1.1 GFI Ratio Trend. We analyzed the monthly GFI ratio over the four-year period (July 2021 to June 2025). A Mann-Kendall trend test indicated a statistically significant decreasing trend ($\tau = -0.44$, $p < 0.001$). However, as shown in Figure 2(a), this decline was not gradual: the yearly average GFI ratio remained stable from Y1 (0.92%) through Y3 (0.88%), then dropped sharply in Y4 (0.57%).

To distinguish whether this decline was gradual or structural, we applied the Pettitt change-point test, which detected a statistically significant structural break at January 2024 ($K = 445$, $p < 0.001$). The mean GFI ratio before the change point (July 2021–December 2023) was 0.95%, compared to 0.58% after (January 2024–June 2025), representing a 39% decrease. Since the change point falls in the middle of Y3, the yearly average for Y3 (0.88%) masks the onset of the decline in its second half.

4.1.2 Repository Heterogeneity in GFI Trends. Of the 37 repositories, 7 (18.9%) never used GFI labels during the analysis period. We conducted Mann-Kendall trend tests on the remaining 30 repositories. After Benjamini-Hochberg correction for 30 simultaneous tests, these repositories showed substantial heterogeneity in GFI usage trends: 7 (23.3%) showed a decreasing trend, 21 (70.0%) showed no significant trend, and 2 (6.7%) showed an increasing trend. This variation could not be explained by repository age (Spearman's $\rho = -0.105$, $p = 0.588$) or primary programming language.

Comparing characteristics across the three trend groups (Kruskal-Wallis test), no statistically significant differences were found in star count, repository age, total issue count, or GFI count (all $p > 0.2$). These results suggest that GFI usage trends cannot be explained by objective characteristics such as project size, maturity, or primary language, but rather depend heavily on project-specific strategic decisions.

4.1.3 Newcomer Engagement with GFIs. As shown in Figure 2(b), the proportion of GFI issues addressed by newcomers remained stable throughout the period. The overall engagement rate was 27.0% (891 out of 3,300 GFI issues), with per-year rates between 25% and 29%. A Mann-Kendall trend test confirmed no significant trend ($\tau = 0.06$, $p = 0.52$). Despite the sharp decline in GFI ratio observed in Y4 (Figure 2(a)), the stable engagement rate (Figure 2(b)) suggests that GFI labels maintain a consistent level of attractiveness for newcomers regardless of their prevalence.

4.2 RQ2: How do task-type labels relate to merge outcomes, and what factors are associated with merge success?

RQ1 established that GFI availability is declining while newcomer engagement remains stable. We now examine whether newcomers who engage with GFIs succeed in having their contributions merged, and what factors are associated with merge outcomes.

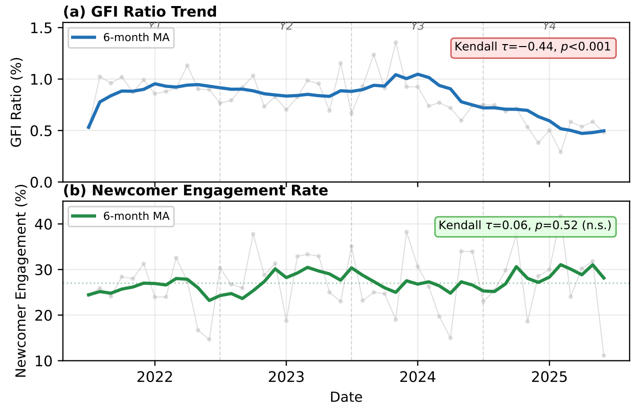

Figure 2: RQ1 time-series trends. (a) Monthly GFI ratio shows a significant decreasing trend ($\tau = -0.44$, $p < 0.001$). (b) Newcomer engagement rate remains stable at approximately 27% ($\tau = 0.06$, n.s.). Gray dots: monthly values; colored lines: 6-month moving averages; dashed vertical lines: analysis year boundaries.

4.2.1 PR Metrics Trends and Task Type Analysis. We analyzed 1,117 GFI-labeled PRs (after excluding bots) spanning all 30 GFI-using repositories (median: 22 PRs/repo, IQR: 5–54), with no single repository exceeding 17% of the total. The overall merge rate was 53.0%. Table 1 summarizes the time-series trends for key metrics. The merge rate showed a significant decreasing trend, while description length showed an increasing trend. The robustness of the aggregate declining trend is further supported by the sensitivity analysis reported in Section 4.2.1.

We analyzed merge rates by task type over four analysis years (Table 2). Bug-fix tasks had the highest overall merge rate (68.7%), peaking at 83.5% in Y2 before declining to 45.9% in Y4. Feature tasks remained stable at approximately 54% with no significant trend. The "Other" category, which includes PRs whose GFI issues lack standard task-type labels (e.g., issues labeled only with module or area tags), showed the steepest decline from 57.1% (Y1) to 28.6% (Y4). This decline was partly driven by a single project (PyTorch), which relies on module-based labels (e.g., `module: autograd`) rather than standard task-type labels (Bug, Feature, Documentation), causing the vast majority of its GFI PRs to be classified as "Other." PyTorch alone contributed 188 PRs to this category with a 0% merge rate. As a sensitivity check, excluding PyTorch's PRs, the "Other" category still declined from 64.4% (Y1) to 48.0% (Y4), confirming that the downward trend is not solely attributable to PyTorch's labeling practices.

4.2.2 Factors Associated with Merge Success. Table 3 compares merged and unmerged GFI PRs. Among initial PR

Table 1: Time-series trend analysis of GFI metrics (Mann-Kendall test)

Metric	Y1	Y4	Kendall τ	Trend
GFI Ratio (%)	0.92	0.57	-0.44***	Decreasing
Newcomer Engagement (%)	25.1	27.7	0.06	No trend
Merge Rate (%)	61.9	42.2	-0.35**	Decreasing
Description Length	306	481	0.35**	Increasing

Note: $^{**}p < 0.01$, $^{***}p < 0.001$ (Holm-Bonferroni adjusted within each table). Monthly Mann-Kendall test over 48 months. Y1/Y4 are yearly averages of monthly values.

Table 2: Merge rate by task type and analysis year

Task Type	Y1	Y2	Y3	Y4	Total	Trend
Bug	64.5%	83.5%	71.9%	45.9%	68.7%	Decr.*
Feature	53.7%	54.8%	53.3%	55.6%	54.4%	None
Docs	68.4%	65.2%	42.4%	47.7%	52.9%	Decr.*
Other	57.1%	46.4%	33.3%	28.6%	40.7%	Decr.*

Note: Y1=Jul'21–Jun'22, Y2=Jul'22–Jun'23, Y3=Jul'23–Jun'24, Y4=Jul'24–Jun'25.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ (Mann-Kendall test, Holm-Bonferroni adjusted).

Table 3: GFI PR metrics by merge status

Metric	Merged	Not Merged	p-value	$\|r\|$
Initial PR characteristics				
Insertions (log)	3.02	2.89	0.885	–
Deletions (log)	1.10	1.39	0.994	–
Changed Files	2.0	2.0	0.155	–
Description Length	382.5	435.0	0.608	–
Process-level metrics†				
Commits Count	3.0	2.0	<0.01**	0.15
Review Count	2.0	1.0	<0.001***	0.32

Note: $^{**}p < 0.01$, $^{***}p < 0.001$ (Holm-Bonferroni adjusted). Median values shown. Mann-Whitney U test. $|r|$: rank-biserial effect size.
†Process-level metrics accumulate during the review lifecycle and are subject to mechanical confounds (see text).

characteristics—code size, number of changed files, and description length—none showed a statistically significant association with merge outcomes. The two process-level metrics (commit count and review count) showed statistically significant differences, but these metrics accumulate during the review lifecycle and are subject to confounds discussed in Section 5.2.

5 DISCUSSION

5.1 Interpretation of RQ1 Findings

The GFI ratio remained stable from Y1 (0.92%) through Y3 (0.88%) before declining sharply in Y4 (0.57%). Crucially, the decline is not gradual: the Pettitt change-point test locates a structural break at January 2024, indicating that the shift began in the second half of Y3 and accelerated through Y4. This trend also varied substantially across repositories—after Benjamini-Hochberg correction, only 23.3% showed a statistically significant decrease, while 70.0% showed no significant trend—and could not be explained by repository age or programming language.

The decline in GFI ratio is both broad and uneven. Among the 30 GFI-using repositories, 77% (23/30) exhibited a decrease in GFI ratio (regardless of statistical significance). The top five declining repositories accounted for 61% of the total GFI decrease, while the remaining 25 repositories contributed 39%. This pattern suggests a compound dynamic: a few projects made substantial cuts while many others experienced moderate declines.

These results suggest that changes in GFI usage are not uniformly driven by a single external factor, but are strongly dependent on project-specific strategic decisions. Some projects may be reducing their use of GFI labels due to changes in their issue triage process, constraints on maintenance resources, or shifts in their community growth strategy. Conversely, projects actively seeking to attract new contributors may be increasing their use of GFIs.

Notably, the proportion of GFIs addressed by newcomers remained stable at approximately 27% throughout the four-year period, with no significant trend. Despite the GFI ratio decline in Y4, this stable engagement rate suggests that GFI labels maintain a consistent level of attractiveness for newcomers. However, 73% of GFIs did not receive a newcomer PR, consistent with Tan et al. [6]'s finding that many GFIs remain unaddressed by newcomers.

5.2 Interpretation of RQ2 Findings

The merge rate for newcomer GFI PRs showed a decreasing trend across all task types except Feature, which remained stable. Bug-fix tasks maintained the highest merge rate (68.7%), likely because their well-defined scope and clear acceptance criteria make them more amenable to newcomer contributions. The steep decline in the "Other" category (57.1% to 28.6%) is partly an artifact of project-specific labeling practices—as shown in Section 4.2.1, a single project's (PyTorch's) module-based labels accounted for much of this decline. Nonetheless, the overall declining trend in merge rates is not reducible to this single-project effect: Bug-fix tasks also declined sharply (83.5% in Y2 to 45.9% in Y4), and the overall merge rate decrease is observed across multiple task types and projects. This confirms that the decline reflects a genuine shift in GFI effectiveness rather than a labeling artifact, while underscoring the importance of accounting for project heterogeneity in aggregate analyses.

On the other hand, the description length of newcomer PRs increased significantly over the same period. This trend may suggest that project quality standards have risen, or that newcomers have learned to provide more detailed descriptions. However, the fact that the merge rate has decreased despite the increase in description length indicates that detailed descriptions alone are not a decisive factor for merge success.

Regarding merge success factors, the quantitative characteristics of the initial submission—code size, number of changed files, and description length—showed no significant association with merge outcomes. What determines merge success remains an open question from our data alone, though we note that the observed direction for description length (unmerged PRs had a 14% higher median than merged PRs) warrants caution against interpreting non-significance as evidence of equivalence.

The process-level metrics (commit count and review count) showed statistically significant differences, but should not be interpreted as independent predictors. Review count in particular is prone to structural confounds: branch protection rules and selective maintainer attention may both inflate review counts for merged PRs irrespective of PR quality. We therefore do not treat these metrics as predictors of merge success.

The parallel declines in GFI supply (RQ1) and merge rate (RQ2) call for interpretation, though the observational nature of our data precludes causal conclusions. We identify three plausible explanations, each with distinct implications:

(H1) Reduced maintainer capacity. Projects reducing GFI labeling may simultaneously allocate fewer resources to reviewing newcomer PRs, causing both metrics to decline together. This is consistent with Tan et al.'s finding that mentoring support—not task recommendation alone—is the key driver of newcomer success [14].

(H2) Rising quality standards. As projects mature, maintainers may apply stricter acceptance criteria, making it harder for newcomer PRs to be merged regardless of GFI availability.

(H3) Shifting task landscape due to generative AI. The structural break in January 2024 coincides with the widespread adoption of LLM-based coding tools. If AI assistants are increasingly handling simple, self-contained tasks, the pool of tasks that are both accessible to newcomers and non-trivially valuable to maintainers may be shrinking—reducing the incentive to label GFIs and the likelihood of accepting AI-assisted submissions.

These hypotheses are not mutually exclusive, and distinguishing among them requires further investigation.

5.3 Implications for Practice

For Project Maintainers: Our findings indicate that GFI labels continue to attract a stable proportion of newcomers (approximately 27% engagement rate throughout the four-year period). As both the GFI ratio and newcomer PR merge rate show declining trends, actively creating and labeling GFIs—and sustaining review support for newcomer PRs—can provide a competitive advantage in newcomer acquisition. Notably, bug-fix tasks have the highest merge rate (68.7%), suggesting that maintainers should prioritize labeling well-scoped bug fixes as GFIs. Regardless of the underlying cause (reduced capacity, higher quality standards, or AI-driven task displacement), the stable newcomer demand implies that investment in GFI labeling and review continues to yield returns in community growth.

For Newcomers: A key finding is that despite newcomers submitting increasingly detailed PRs over the study period (description length increased by 47%), merge rates continued to decline. The quantitative characteristics of the initial submission do not predict merge outcomes, leaving open the question of what does. Prior work has found that maintainer mentoring—not the initial submission quality—is the primary driver of newcomer success [14]; newcomers should therefore seek out projects and maintainers that actively support contributions, and proactively engage with reviewer feedback once a PR is submitted. GFI aggregators [16] can help identify suitable tasks efficiently.

For Researchers: This study emphasizes the importance of considering project heterogeneity in OSS onboarding research. Analyses based solely on aggregated statistics may overlook the diverse strategies adopted by individual projects. Additionally, our label-based task-type analysis demonstrates that more granular categorization can yield actionable insights. However, researchers should be aware that labeling practices vary significantly across projects—some use task-type labels (bug, feature, documentation), while others use module-based or area-based labels that do not fit standard classification schemes. In our dataset, 4.2% of PRs matched multiple task types, requiring manual classification by distinguishing type labels from area labels. This heterogeneity should be accounted for in cross-project studies.

5.4 Threats to Validity

Construct Validity: In this study, we used the list of GFI labels presented by Tan et al. [6] and Turzo et al. [8]. However, some projects may adopt their own label naming conventions, which means we may not have captured all GFIs. Our task-type classification (Bug, Feature, Documentation, Other) is based on issue label keywords, with 47 overlapping cases (4.2%) manually classified by the first author; this single-rater process lacks inter-rater reliability verification, though the small proportion limits its impact on overall results. Some projects use alternative labeling schemes (e.g., module-based labels) that do not map cleanly to these categories. Additionally, we defined newcomers as 'individuals submitting their first PR to a repository,' which does not take into account their overall experience on GitHub. Our bot detection relies on GitHub's author_type field, which may not capture all automated contributions (e.g., bots configured as regular users or semi-automated tools used by human developers).

Internal Validity: This study observes time-series trends and does not claim specific causal relationships. The observed changes may be influenced by multiple confounding factors, including overall changes in the OSS ecosystem, strategic decisions by individual projects, and the evolution of development tools. Because our GFI label data is a point-in-time snapshot collected in November 2025, one might suspect that the sharp Y4 decline in GFI ratio reflects labeling lag rather than a genuine trend. However, we confirmed that repositories with reduced GFI counts in Y4 continued to create issues at their usual volume (e.g., hundreds of issues per month) yet assigned zero GFI labels for periods of seven to nine months—well beyond plausible triage delays—indicating a real change in labeling practice rather than a data-collection artifact. Additionally, because our snapshot captures only the labels present at collection time, labels added after issue creation or later removed are not distinguished. We verified via the GitHub Timeline API on a stratified sample of 100 GFI issues that 92% received their GFI label within 7 days of creation, confirming that the label-creation date is a reasonable proxy for the labeling decision. In RQ2, 95 GFI issues (9.7%) received multiple newcomer PRs; we treat each PR as an independent contribution attempt, which may introduce non-independence among observations linked to the same issue.

External Validity: This study is limited to popular OSS projects (top-starred) on GitHub. Different trends may be observed in smaller projects or on other platforms (e.g., GitLab, Bitbucket). Furthermore, the analysis is confined to software projects, excluding non-software repositories such as documentation and learning resources.

6 CONCLUSION

We analyzed 406,826 issues (identifying 3,300 GFI-labeled issues), 43,906 newcomer pull requests, and 1,117 GFI PRs from 37 popular OSS projects on GitHub to investigate the GFI ratio, newcomer engagement patterns, and the changing characteristics of GFI PRs over a four-year period from July 2021 to June 2025.

Regarding RQ1, a change-point analysis identified a structural break in January 2024, after which the GFI ratio declined significantly. This trend varied greatly among repositories, driven by project-specific decisions rather than observable project characteristics. Newcomer engagement remained stable at approximately 27%, indicating sustained demand despite declining supply.

Regarding RQ2, examining the outcomes of newcomer engagement, the merge rate also declined over the period. Bugfix tasks maintained the highest merge rate (68.7%), while initial PR characteristics showed no association with merge outcomes, suggesting that GFIs remain appropriately scoped. The steep decline in the "Other" category (57.1% to 28.6%) was partly attributable to project-specific labeling practices, highlighting the importance of accounting for heterogeneity in cross-project analyses.

Together, these findings reveal a widening gap between stable newcomer interest in GFIs and the declining availability and success of GFI-based onboarding, underscoring the need for maintainers to sustain both GFI labeling and review support.

The findings of this study offer practical implications for project maintainers designing strategies for attracting new contributors. Given the decline in GFI ratio observed in Y4, proactively creating and labeling GFIs can provide a competitive advantage for newcomer acquisition. For newcomers, exploring projects that actively maintain GFI labels and engaging proactively with reviewer feedback—consistent with prior evidence on mentoring [14]—remains a practical strategy.

Future research would benefit from (1) a qualitative investigation into the decision-making processes behind why some projects are increasing their use of GFIs while others are decreasing it; (2) identification of the root causes for the decline in GFI PR merge rates (e.g., issue quality, maintenance resources, changes in quality standards); (3) examination of task-type-specific support strategies for newcomers; and (4) development of AI-assisted tools for automatically detecting and recommending GFI candidates, which could reduce the labeling burden on maintainers and help counteract the observed decline in GFI availability. Such studies are expected to deepen our understanding of effective newcomer onboarding strategies in sustainable OSS ecosystems.

DATA AVAILABILITY

The replication package for this study is available at https://doi.org/10.5281/zenodo.17638558.

REFERENCES

[1] I. Steinmacher, T. Conte, M. A. Gerosa, and D. Redmiles, "Social barriers faced by newcomers placing their first contribution in open source software projects," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, ser. CSCW '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1379–1392. [Online]. Available: https://doi.org/10.1145/2675133.2675215

[2] D. Sholler, I. Steinmacher, D. Ford, M. Averick, M. Hoye, and G. Wilson, "Ten simple rules for helping newcomers become contributors to open projects," *PLoS computational biology*, vol. 15, no. 9, p. e1007296, 2019.

[3] I. Steinmacher, A. P. Chaves, T. U. Conte, and M. A. Gerosa, "Preliminary empirical identification of barriers faced by newcomers to open source software projects," in *2014 Brazilian Symposium on Software Engineering*, 2014, pp. 51–60.

[4] W. Xiao, H. He, W. Xu, X. Tan, J. Dong, and M. Zhou, "Recommending good first issues in github oss projects," in *Proceedings of the 44th International Conference on Software Engineering*, ser. ICSE '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 1830–1842. [Online]. Available: https://doi.org/10.1145/3510003.3510196

[5] Y. Huang, J. Wang, S. Wang, Z. Liu, D. Wang, and Q. Wang, "Characterizing and predicting good first issues," in *Proceedings of the 15th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, ser. ESEM '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: https://doi.org/10.1145/3475716.3475789

[6] X. Tan, M. Zhou, and Z. Sun, "A first look at good first issues on github," in *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, ser. ESEC/FSE 2020. New York, NY, USA: Association for Computing Machinery, 2020, p. 398–409. [Online]. Available: https://doi.org/10.1145/3368089.3409746

[7] V. N. Subramanian, I. Rehman, M. Nagappan, and R. G. Kula, "Analyzing first contributions on github: What do newcomers do?" *IEEE Software*, vol. 39, no. 1, pp. 93–101, 2022.

[8] A. K. Turzo, S. Sultana, and A. Bosu, "From first patch to long-term contributor: Evaluating onboarding recommendations for oss newcomers," *IEEE Trans. Softw. Eng.*, vol. 51, no. 4, p. 1303–1318, Apr. 2025. [Online]. Available: https://doi.org/10.1109/TSE.2025.3550881

[9] I. Steinmacher, T. U. Conte, C. Treude, and M. A. Gerosa, "Overcoming open source project entry barriers with a portal for

newcomers," in *Proceedings of the 38th International Conference on Software Engineering*, ser. ICSE '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 273–284. [Online]. Available: https://doi.org/10.1145/2884781.2884806

[10] F. Santos, B. Trinkenreich, J. a. F. Pimentel, I. Wiese, I. Steinmacher, A. Sarma, and M. A. Gerosa, "How to choose a task? mismatches in perspectives of newcomers and existing contributors," in *Proceedings of the 16th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement*, ser. ESEM '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 114–124. [Online]. Available: https://doi.org/10.1145/3544902.3546236

[11] W. Xiao, J. Li, H. He, R. Qiu, and M. Zhou, "Personalized first issue recommender for newcomers in open source projects," in *Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering*, ser. ASE '23. IEEE Press, 2024, p. 800–812. [Online]. Available: https://doi.org/10.1109/ASE56229.2023.00158

[12] I. Steinmacher, S. Balali, B. Trinkenreich, M. Guizani, D. Izquierdo-Cortazar, G. G. Cuevas Zambrano, M. A. Gerosa, and A. Sarma, "Being a mentor in open source projects," *Journal of Internet Services and Applications*, vol. 12, no. 1, p. 7, Sep 2021. [Online]. Available: https://doi.org/10.1186/s13174-021-00140-z

[13] S. Balali, U. Annamalai, H. S. Padala, B. Trinkenreich, M. A. Gerosa, I. Steinmacher, and A. Sarma, "Recommending tasks to newcomers in oss projects: How do mentors handle it?" in *Proceedings of the 16th International Symposium on Open Collaboration*, ser. OpenSym '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3412569.3412571

[14] X. Tan, Y. Chen, H. Wu, M. Zhou, and L. Zhang, "Is it enough to recommend tasks to newcomers? understanding mentoring on good first issues," in *Proceedings of the 45th International Conference on Software Engineering*, ser. ICSE '23. IEEE Press, 2023, p. 653–664. [Online]. Available: https://doi.org/10.1109/ICSE48619.2023.00064

[15] M. Guizani, T. Zimmermann, A. Sarma, and D. Ford, "Attracting and retaining oss contributors with a maintainer dashboard," in *Proceedings of the 2022 ACM/IEEE 44th International Conference on Software Engineering: Software Engineering in Society*, ser. ICSE-SEIS '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 36–40. [Online]. Available: https://doi.org/10.1145/3510458.3513020

[16] S. Gazanchyan, "Awesome first pr opportunities," 2020, [Online]. Available: https://github.com/MunGell/awesome-for-beginners.