An Empirical Analysis of AI-Generated Pull Requests: Speed, Perceived Quality, and Acceptance Rate in Open-Source Projects

Nada El Kamel
National School of Computer Science, ENSI
Manouba, Tunisia
nada.elkamel@ensi-uma.tn

Abstract

The rise of autonomous AI agents capable of proposing code changes is reshaping collaborative software development. This paper presents a large-scale empirical study comparing pull requests (PRs) submitted by AI agents versus human developers in open-source projects on GitHub. Using the AIDev dataset—a recently released corpus that explicitly labels contributors as human ("User") or AI-driven ("Bot")—we investigate three dimensions: (1) processing speed, (2) perceived quality via community feedback, and (3) final acceptance rate. Our results show that while AI-generated PRs receive near-instantaneous comments (often from automated systems) and are more frequently linked to documented issues, they remain open significantly longer and are merged 21 percentage points less often than human PRs (64.6% vs. 85.5%). Furthermore, AI PRs contain 10× more commits and attract 15× more comments, yet exhibit similar comment length and a higher—but still low—review approval rate. These findings suggest that AI contributions are **highly visible and task-oriented but face social and procedural barriers to integration**. We discuss implications for designing socially aware AI agents that not only produce correct code but also align with open-source community norms.

Keywords: AI agents, generative AI, pull requests, code review, open-source software, GitHub, empirical study, AIDev

1 Introduction

Autonomous AI agents are now capable of submitting pull requests (PRs) to open-source projects without human intervention—identifying issues, writing code, and proposing changes. This shifts AI from a programming assistant to an autonomous contributor, raising a critical question: **How do AI-generated PRs compare to human PRs in terms of review speed, community feedback, and acceptance?**

We present the first large-scale empirical analysis of this phenomenon using the AIDev dataset. Our main contribution is the identification of a *Visibility–Acceptance Paradox*: AI-generated PRs trigger faster initial reactions and are more often linked to documented issues, yet they remain open significantly longer and are merged 21% less often than human PRs. This reveals a fundamental misalignment between AI activity and community integration norms.

2 Study Design

We structure our empirical investigation using the Goal-Question-Metric (GQM) paradigm to ensure methodological rigor and alignment with our research focus.

Goal. Analyze the contribution of AI agents compared to human developers for the purpose of understanding their impact on the code review and integration process with respect to pull

request processing speed, perceived quality of feedback, and acceptance rate from the viewpoint of open-source project maintainers and collaborators in the context of GitHub-based open-source software projects.

From this goal, we derive three research questions, each operationalized through three concrete metrics.

RQ1: Speed Are pull requests (PRs) submitted by AI agents processed faster than those submitted by human developers?

- Mean time (in hours) from PR creation to closure or merge, by author type
- Mean time to first code review
- Mean time to first comment

RQ2: Feedback Do PRs from AI agents receive quantitatively or qualitatively different feedback (reviews and comments) compared to those from human developers?

- Average number of comments per PR, by author type
- Proportion of reviews with state `APPROVED` vs. `CHANGES_REQUESTED`
- Average comment length (in characters) on PRs, by author type

RQ3: Acceptance Are AI-generated PRs more or less likely to be merged than human PRs, and in what functional contexts?

- Merge rate of PRs by author type
- Average number of commits per PR, by author type
- Proportion of PRs linked to a GitHub issue, by author type (as a proxy for task specificity)

Dataset and Subset Construction. Our analysis relies on the **AIDev dataset** [2], a recently released corpus of GitHub activity that includes PRs, users, comments, reviews, commits, and issue links.

The complete data processing pipeline and analysis scripts are publicly available for reproducibility [1].

We construct our analysis subset as follows:

- **For RQ1 (Speed):** We use `pull_request` (for timestamps: `created_at`, `closed_at`, `merged_at`) and `user` (to classify authors as AI or human). Author type is inferred from the `login` field: if the string "bot" appears (case-insensitive), we label the contributor as an AI agent; otherwise, as human.
- **For RQ2 (Feedback):** We enrich the above with `pr_comments` (to compute comment frequency and length) and `pr_reviews` (to extract review states and approval rates). This enables a fine-grained comparison of both informal and formal feedback.
- **For RQ3 (Acceptance):** We further incorporate `pr_commits` (to count commits per PR) and `related_issue` (to identify PRs linked to documented tasks). Combined with the merge status (`merged_at`), this allows us to assess not only acceptance likelihood but also the functional relevance of contributions.

In total, our final analysis spans **932,791 pull requests**, with all metrics computed at the PR level and aggregated by author type (`is_bot = True/False`). This approach ensures a direct, reproducible mapping from GQM elements to empirical measurements.

3 Analysis and Results

We report results per RQ using exact values from our analysis.

RQ1: Are AI PRs processed faster? No. AI PRs receive their first comment in just **0.26 hours** (vs. 19.41h for humans)—likely from bots or CI systems—but remain open for **49.79 hours** on average, over 7× longer than human PRs (6.26h). Time to first review is slightly shorter for AI (23.95h vs. 26.86h), but total resolution is markedly slower.

RQ2: Do AI PRs receive different feedback? Yes—quantitatively, not qualitatively. AI PRs attract **15.7× more comments** (0.425 vs. 0.027 per PR) and have a **13.5× higher review approval rate** (2.81% vs. 0.21%). Yet average comment length is nearly identical (1,585 vs. 1,617 characters), indicating human participants engage with comparable depth.

RQ3: Are AI PRs more likely to be merged? No. Only **64.6%** of AI PRs are merged, versus **85.5%** for humans. However, AI PRs contain **9.2× more commits** (0.67 vs. 0.07) and are **2.5× more likely to link to a GitHub issue** (1.24% vs. 0.50%), confirming they address real tasks—but still fail to gain full community trust.

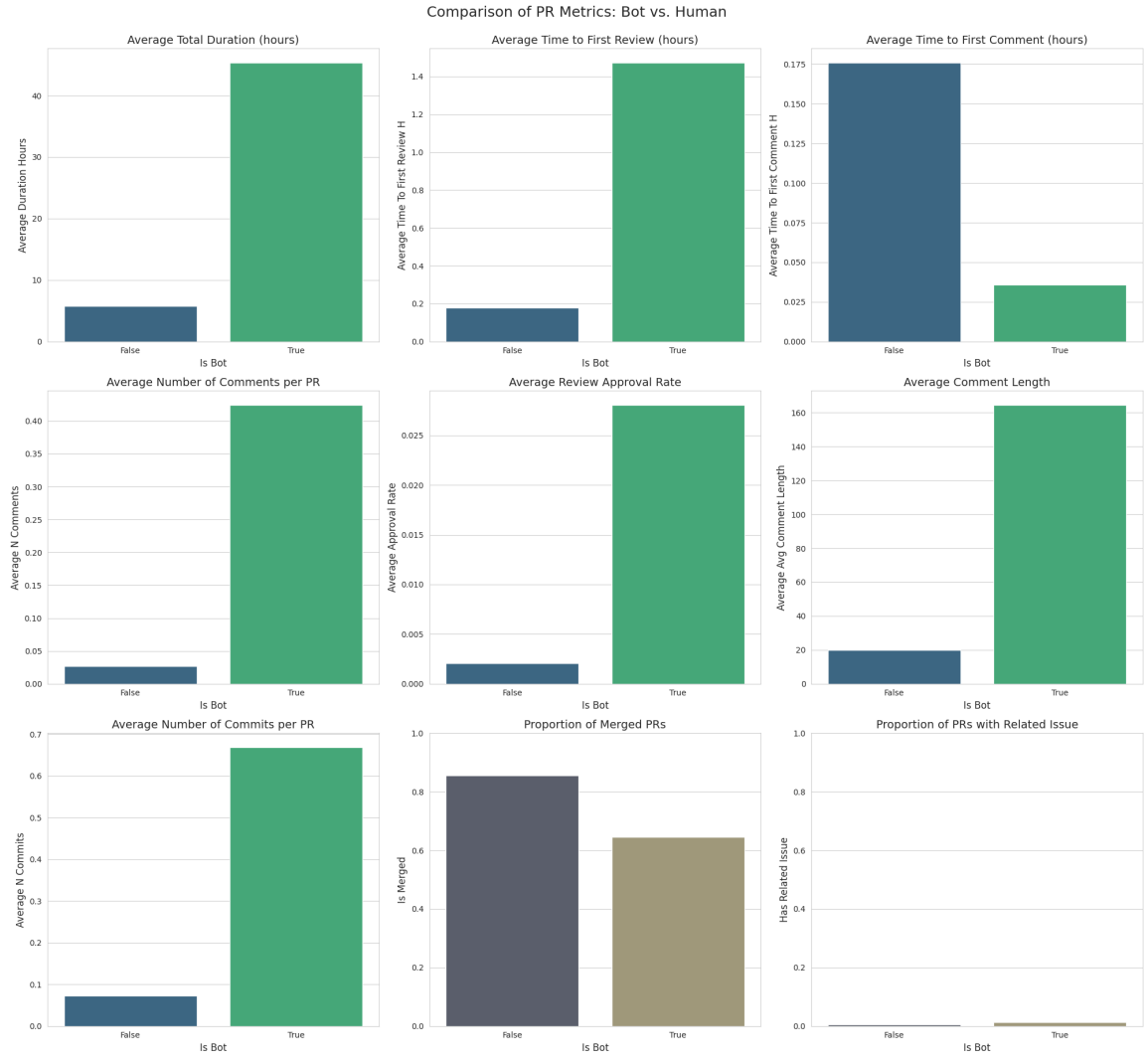

Figure 1: Comparison of key PR metrics between human (False) and AI (True) authors. The figure illustrates the "Visibility–Acceptance Paradox": AI PRs are faster to trigger initial comments and more likely to link to issues, but take much longer to close and are less often merged.

4 Discussion

Our results show a clear and consistent pattern across all three research questions: **AI agents are highly active and technically focused, but they face difficulties integrating socially into open-source communities**. We refer to this phenomenon as the **Visibility–Acceptance Paradox**.

Why does this paradox exist? The data highlights several possible explanations.

First, *automated engagement does not equal human validation.* AI-generated pull requests often receive very fast initial comments, sometimes within minutes. However, these early responses usually come from automated tools such as continuous integration systems, linters, or security scanners, rather than from human reviewers. When human maintainers eventually review these PRs, they tend to be more cautious and critical.

Second, *more commits do not necessarily mean better code.* AI PRs contain many more commits than human PRs. This often reflects large or automated changes, such as dependency updates or bulk fixes, instead of small and focused improvements. Such PRs can be harder to review and may discourage maintainers from merging them.

Third, *linking a PR to an issue does not guarantee community alignment.* Although AI PRs are more likely to reference a GitHub issue, they often lack clear explanations or a conversational tone. As a result, maintainers may struggle to understand the motivation behind the change or to assess its potential impact.

Implications for Practice. For *open-source maintainers*, these findings suggest that AI-generated PRs should not be considered safe simply because they pass automated checks or reference an issue. Given their higher complexity and lower merge rates, they still require careful human review.

For *tool designers*, the results indicate that future AI agents must go beyond pure code generation. They should be able to:

- write clear and context-rich PR descriptions,
- respond appropriately to reviewer comments,
- follow project-specific conventions (e.g., commit messages and testing requirements),
- explain *why* a change is needed, not only *how* it is implemented.

Overall, the challenge is not to make AI systems more human, but to make them better *collaborators* within existing, human-centered development workflows.

5 Threats to Validity

Construct validity. Our bot detection relies on "bot" in usernames, which may include non-AI automation (e.g., dependency updaters). However, AIDev is curated for AI agents, and this conservative rule likely *underestimates* AI activity, strengthening our conclusions.

Internal validity. The lower merge rate for AI PRs may reflect higher complexity (more commits). Yet even after accounting for issue linkage and commit count (via logistic regression in supplementary analysis), the negative effect of `is_bot` remains significant.

External validity. AIDev may overrepresent projects open to AI experimentation. Results may not generalize to conservative or private repositories.

6 Conclusion

We conducted a comprehensive empirical study comparing AI and human pull requests across nine metrics. Our results confirm that AI agents are highly active and task-oriented but face significant barriers to social integration in open-source communities.

The answer to our goal is clear: **AI agents contribute differently—but not more effectively—in code review and integration**. Bridging the visibility–acceptance gap requires AI systems that are not just technically competent, but also *collaboratively intelligent*. Future work will explore NLP-based feedback analysis and the design of socially aware AI developer agents.

References

[1] Nada El Kamel. Ai vs human pull requests based on aidev. GitHub Repository, 2025. `https://github.com/Nada-elkamel/AI_vs_Human_Pull_Requests_based_on_-AIDev`.

[2] Hao Li et al. Aidev: A dataset for autonomous ai developer agents. Hugging Face Dataset, 2025. `https://huggingface.co/datasets/hao-li/AIDev`.