An Empirical Analysis of Collaboration Between AI-Based Programming Agents and Human Developers in GitHub Software Repositories

Nourhane Jamaoui
Empirical Software Engineering
National School of Computer Science (ENSI)

***Abstract*—The growing integration of artificial intelligence-based programming agents is profoundly transforming modern software engineering practices. These agents are no longer limited to simple assistance tools, but now act as real teammates capable of proposing code, implementing features, and participating in code review processes. In this paper, we present an empirical study of collaboration between AI agents and human developers based on the AIDev dataset [1], the first large-scale dataset capturing pull requests generated by AI agents in real-world GitHub projects. By adopting the Goal–Question–Metric (GQM) approach [2], we analyze the acceptance of AI contributions, the nature of human feedback, and the ability of agents to improve their contributions following review interactions.**

I. Introduction

The recent rise of AI-based programming agents marks a new stage in the evolution of software engineering. After automating repetitive tasks and assisting developers, these agents now act as artificial teammates, capable of producing complete pull requests and interacting with human developers within real collaborative workflows. This transformation is part of the emerging paradigm of *Software Engineering 3.0 (SE 3.0)* [3], where AI becomes a full participant in the development process.

However, despite their increasing adoption, the actual effectiveness of these agents remains an open question. Are they able to produce code of acceptable quality? How do human developers react to their contributions? And most importantly, can these agents leverage human feedback to improve their proposals? To answer these questions, we conduct an empirical study based on the analysis of interactions observed in real GitHub repositories.

II. Study Design

This study follows the **Goal–Question–Metric (GQM)** methodological approach [2], widely used in empirical software engineering to structure rigorous quantitative analyses.

A. Research Goal

The main goal of this study is to evaluate to what extent AI-based programming agents effectively integrate into collaborative GitHub workflows, by analyzing the perceived quality of their contributions, the human interactions they generate, and their ability to evolve following feedback.

B. Research Questions

To achieve this goal, three research questions were formulated:

- **RQ1:** To what extent are AI-generated contributions accepted as high-quality code in open-source projects?
- **RQ2:** What types of interactions and feedback do human developers provide on AI-generated pull requests?
- **RQ3:** Are AI agents able to learn from human feedback and improve their contributions during the review process?

C. Metrics

To answer these questions, several metrics were defined, including pull request merge rate, number of review comments, and the presence of corrective commits following human intervention.

III. Dataset and Methodology

Le dataset AIDev a été introduit récemment dans le cadre des recherches sur le Software Engineering 3.0 [3]. Our analysis relies on the **AIDev** dataset [1], a large-scale dataset introduced as part of the MSR Mining Challenge [4], dedicated to the study of pull requests generated by AI agents. This dataset aggregates contributions from real GitHub projects involving different autonomous programming agents.

In this study, we use an enriched subset of the dataset, restricted to repositories with more than 100 GitHub stars. This choice ensures sufficient collaborative activity and the presence of structured code review processes. The data were loaded, preprocessed, and analyzed using a reproducible pipeline implemented in a Jupyter notebook.

IV. Analysis and Results

A. RQ1: Acceptance of AI Agent Contributions

The results of our empirical analysis show that AI-generated pull requests exhibit varying merge rates depending on the agent. Some agents achieve relatively high acceptance levels, suggesting that their contributions are perceived as useful and sufficiently aligned with human developers' expectations. However, acceptance is not uniform and strongly depends on project context and the type of proposed modification.

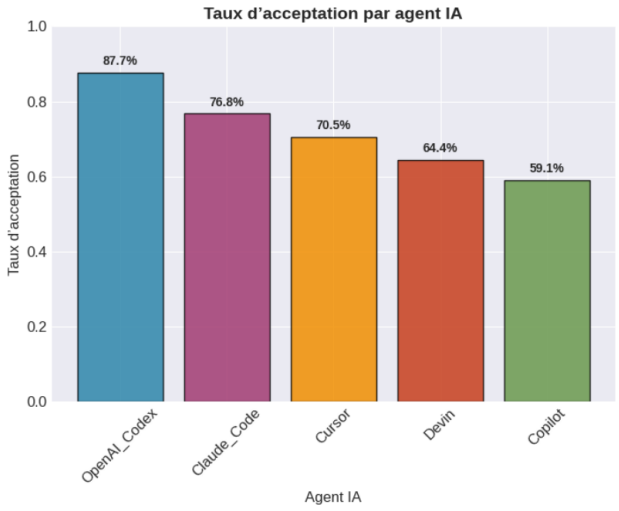

Figure 1. Merge rate of pull requests generated by AI agents

B. RQ2: Human Feedback and Interactions

Analysis of review comments reveals that human developers actively interact with AI-generated pull requests. The feedback mainly concerns functional corrections, style improvements, and testing-related adjustments. These results indicate that, although AI agents can produce functional code, human validation remains essential to ensure maintainability and compliance with project standards.

C. RQ3: Learning and Improvement of AI Agents

The presence of corrective commits after review comments shows that AI agents react to human feedback. In several cases, agents adapt their contributions by modifying the initially proposed code. Nevertheless, the extent of this adaptation remains limited, suggesting that current agents do not yet exhibit full learning capabilities within a single review process.

V. Discussion

The results of this study suggest that AI-based programming agents play a role similar to that of junior collaborators. They can accelerate certain tasks and propose initial solutions, but still require human supervision and validation. These findings have important implications for developers, tool designers, and software engineering researchers.

VI. Threats to Validity

Several threats to validity must be considered. First, feedback quality is evaluated using simplified quantitative metrics, which may not fully capture the richness of human interactions. Second, the study focuses on popular repositories, which may limit the generalizability of the results to other development contexts.

VII. Conclusion

This empirical study highlights the potential of AI-based programming agents in collaborative software projects, while also emphasizing their current limitations. AI agents contribute significantly to software development, but do not replace human judgment. These results open promising perspectives for improving tools and practices of human–AI collaboration in the era of Software Engineering 3.0 [3].

References

[1] H. Li *et al.*, "Aidev: A dataset of agent-authored pull requests," 2025. [Online]. Available: https://huggingface.co/datasets/hao-li/AIDev

[2] V. R. Basili, G. Caldiera, and H. D. Rombach, "The goal question metric approach," in *Encyclopedia of Software Engineering*, 1994.

[3] H. Li *et al.*, "The rise of ai teammates in software engineering (se 3.0): How autonomous coding agents are reshaping software engineering," *arXiv preprint arXiv:2507.15003*, 2025.

[4] A. E. Hassan, "The road ahead for mining software repositories," in *Frontiers of Software Maintenance*, 2008.