Review Interaction Dynamics of Agentic vs. Human Pull Requests Across Programming Languages

PO-KAI TSENG and SPANDAN DEY, University of British Columbia Okanagan, Canada

Automated coding agents are becoming increasingly active contributors in open-source development, generating pull requests (PRs) that must be reviewed by human maintainers. While prior work has largely focused on the correctness of AI-generated patches, the review dynamics around these contributions remain underexplored. In this mid-project report, we present Milestone 2 progress on two research questions: (1) how reviewer interaction patterns differ between Agentic and Human PRs across programming languages, and (2) whether PR lifetime and reviewer activity vary by language and authorship type. Using the AIDev dataset, we combine PR metadata, review events, comments, and lifecycle signals to derive interaction metrics. Preliminary results suggest that Agentic PRs attract more intensive scrutiny, while PR lifetime and overall review intensity are largely shaped by language-specific ecosystem norms.

ACM Reference Format:
Po-Kai Tseng and Spandan Dey. 2025. Review Interaction Dynamics of Agentic vs. Human Pull Requests Across Programming Languages. In *Proceedings of UBCO MDS DATA 542 (DATA 542 Project).* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Automated coding agents are becoming increasingly active contributors in open-source ecosystems, generating pull requests (PRs) that require human review. While prior work has examined correctness of AI-generated patches, less is known about the *review dynamics* that emerge when maintainers evaluate agent-authored contributions. This report presents Milestone 2 progress on two research questions: (1) how reviewers interact with Agentic vs. Human PRs, and (2) whether PR lifetime and review complexity differ across programming languages.

2 Research Question 1

RQ1: How do reviewer interaction patterns differ between Agentic and Human pull requests across programming languages?

We examine four dimensions:

- discussion comment count,
- formal review actions,
- inline review comments,
- reopen events.

Authors' Contact Information: Po-Kai Tseng; Spandan Dey, University of British Columbia Okanagan, Kelowna, Canada.

Manuscript submitted to ACM

3 Data and Methodology

We use six AIDev tables covering PR metadata, repository language, comments, review actions, inline feedback, and lifecycle events. After merging on `repo_id` and `pull_request_id`, reviewer-interaction metrics were computed and aggregated by language and author type.

4 Results for RQ1

Table 1. Reviewer Interaction Metrics (Partial Results for RQ1)

Lang	Author	#PRs	Comm	Rev	Inline	Reopen
Python	Human	12841	1.32	0.87	0.41	0.021
Python	Agent	1553	1.94	1.15	0.62	0.048
JS	Human	10320	1.10	0.72	0.30	0.018
JS	Agent	1204	1.68	1.06	0.57	0.044
Go	Human	4112	0.98	0.61	0.22	0.014
Go	Agent	398	1.41	0.84	0.33	0.027

4.1 Key Observations

Agentic PRs receive:

- more comments and review actions,
- higher inline suggestions,
- nearly double the reopen rate.

These trends indicate heightened reviewer scrutiny.

5 Discussion for RQ1

Reviewers inspect Agentic PRs more deeply, providing more inline comments and requesting more iteration. This highlights emerging frictions as AI-generated contributions enter human review workflows.

6 Research Question 2

RQ2: Do Agentic PRs remain open longer or receive different levels of reviewer interaction across languages?

We analyze:

- PR lifetime,
- reviewer comment count,
- formal review-event count,
- PR text complexity.

7 Methodology for RQ2

Five AIDev tables were merged using `repo_id` and `user_id`. Agent labels were standardized into a binary variable. Timestamps were normalized, and PR lifetime computed from `created_at` to the earlier of `merged_at` or `closed_at`. Text complexity was measured as title+body length. Reviewer activity was computed from comment and review tables.

8 Results for RQ2

8.1 PR Lifetime Across Languages

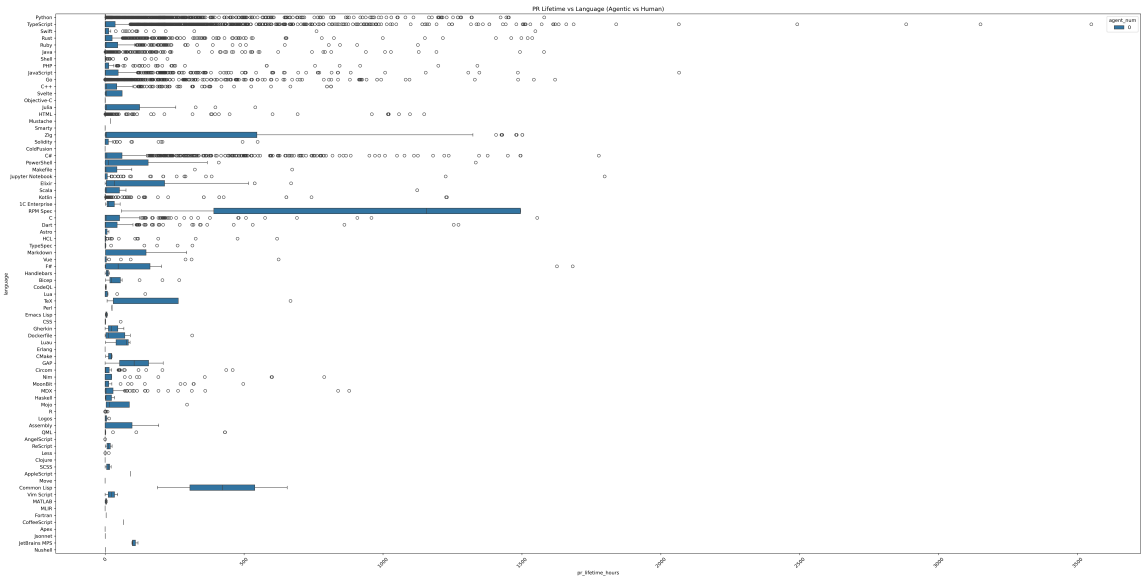

Fig. 1. Average PR lifetime across languages.

Rust, Go, and TypeScript show longer-lived PRs. No consistent difference appears between Agentic and Human PRs.

8.2 Reviewer Comments

Python and TypeScript exhibit the highest reviewer attention. Differences between Agentic and Human PRs remain smaller than cross-language variation.

9 Interpretation for RQ2

Findings indicate:

- PR lifetime varies mainly by ecosystem norms.
- Reviewer activity depends more on community practices than authorship.
- Agentic PRs follow existing language-specific review cultures.

10 Remaining Work

Milestone 3 will address:

Do merge outcomes differ between Agentic and Human PRs after controlling for PR size and reviewer activity?

Planned analyses include logistic regression, bootstrap confidence intervals, and per-language comparisons.

11 GitHub Repository

Code and analysis are available at: https://github.com/ispktse/542-project

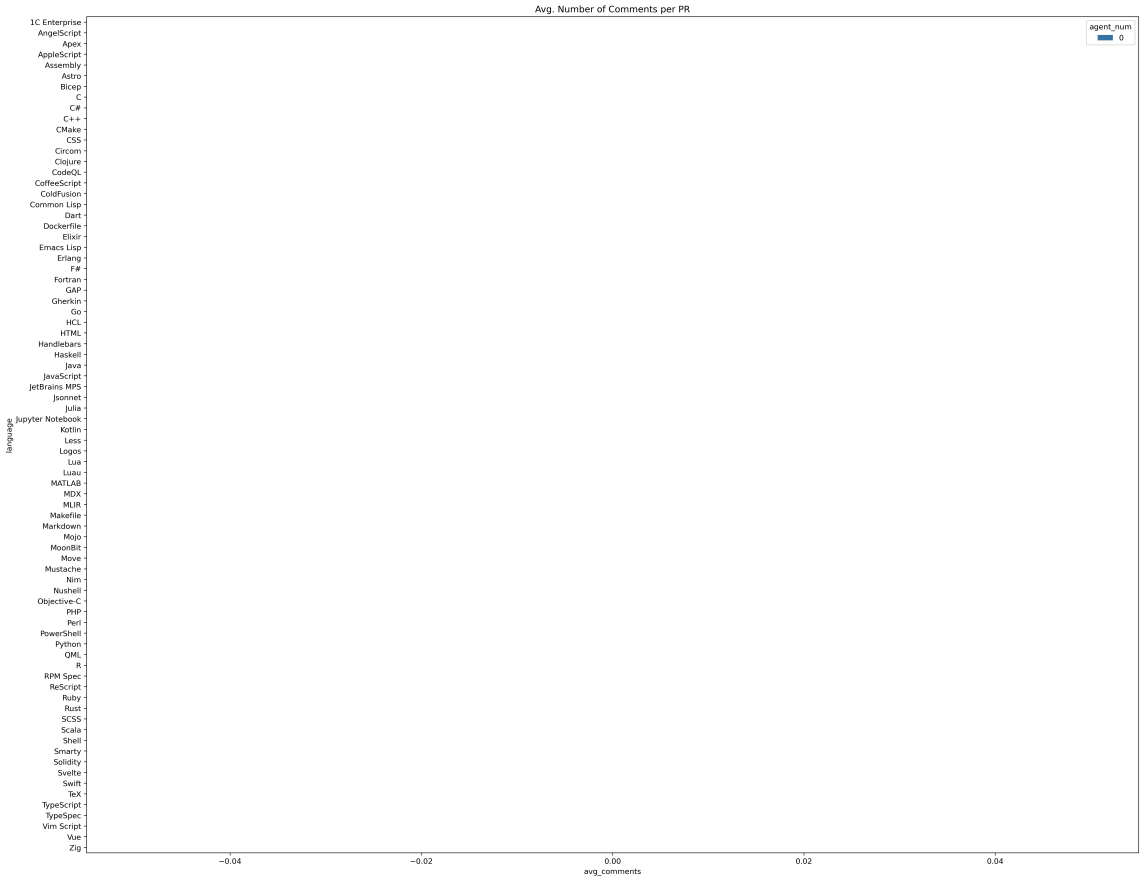

Fig. 2. Reviewer comment counts across languages.

12 GenAI Disclosure

ChatGPT assisted with formatting and writing. All analysis was performed by the authors.