Mouna Tounakti

Acceptance and Trends of AI-Authored
Pull-Requests on GitHub Activity

PART 1

1.Study and Analysis of the Dataset

The AIDev dataset is a relational**,** centered around GitHub pull requests authored by AI agents. It includes the following key semantic layers:

- **Metadata**
  - Pull Requests (titles, descriptions, agent identity, state (open/merged/closed), timestamps)
  - Repositories
  - Users

- **Comments & Reviews**
  - PR comments
  - PR reviews and review comments

- **Comments & Diffs**
  - PR commits and commit diffs

- **Issues & Events**
  - PR timelines
  - Linked issues and issue metadata
  - Automated PR task-type annotations

2. Study Design Using the GQM (Goal–Question–Metric) Approach

2.1 Goal
The goal of this project is to analyze the impact and adoption of AI-generated contributions in open-source software development by examining AI-authored pull requests on GitHub.

2.2 Questions

1. Are AI-authored pull requests more likely to be merged or rejected?

2. How has the volume of AI-authored PRs changed over time?

Mouna Tounakti

PS: *The second question "How do developers respond to pull requests created by AI agents?"; but the dataset does not contain comments or reviews columns, so we cannot directly count comments, reviews with this dataset version. That's why I choose to change this question and its metrics.*

2.3 Metrics

Question 1: Are AI-authored pull requests more likely to be merged or rejected?

- % of PRs that are merged

- % of PRs closed without being merged

Question 2: How has the volume of AI-authored PRs changed over time?

- Monthly count of AI-authored pull requests.
- Linear trend coefficient estimating growth over time.

3. Dataset

-The study uses the Hugging Face release of the AIDev dataset, which consists of pull request–level data from real-world GitHub repositories. Each record corresponds to a single pull request and includes metadata such as creation time, closure or merge time, repository identifier, and AI agent attribution.

- The dataset was filtered to include only AI-authored pull requests, identified by a non-null agent column. Additionally, only pull requests with a final outcome (merged or closed) were retained.

PART 2 – Execution Documentation

1. **Data Loading and Preparation**
   - Load AIDev dataset directly from Hugging Face using the datasets library.
   - Print basic information about the dataset,to understand its structure.
   - Convert dataset into a pandas DataFrame to enable further analysis and preprocessing.
2. **Question 1 steps**
- Identify AI-authored pull requests using the **agent** column (non-null values indicate AI authorship).
- Filter the data to retain only pull requests with a final outcome (merged or closed).
- Create a derived variable:
   - is_merged: indicates whether the pull request was successfully merged.
- Compute the proportion of merged versus non-merged AI-authored pull requests.
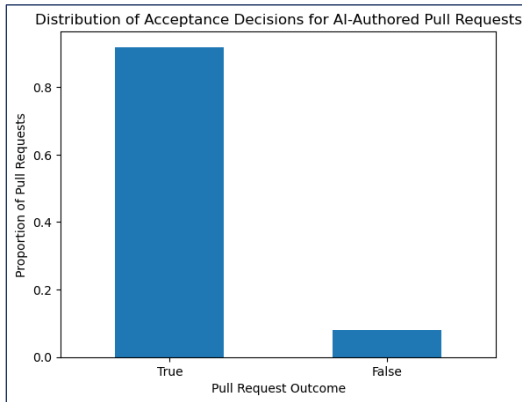- Visualize merge outcomes using a bar chart.

3. **Question 2 steps**

- Use the **created_at** column to analyze pull request creation dates.
- Create a derived time variable:
   - year_month: groups pull requests by year and month.
- Aggregate the number of AI-authored pull requests per month.
- Visualize monthly trends using line plots.

- Apply linear regression model to confirm whether AI-authored pull request activity is increasing over time.
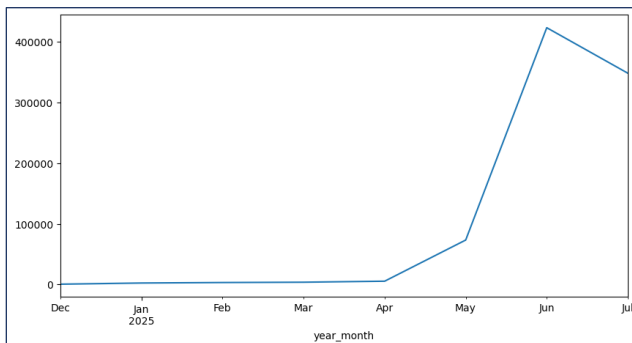
4. **Interpretations**

Question 1: Are AI-authored pull requests more likely to be merged or rejected?

Distribution of Acceptance Decisions for AI-Authored Pull Requests

The results show that **91.88%** of AI-authored pull requests were **successfully merged**, while only **8.12%** were **closed without being merged**.

These findings indicate a high level of acceptance of AI-generated contributions by human developers.

Question 2: How has the volume of AI-authored PRs changed over time?

The results show a **clear and consistent growth** trend in the volume **of AI-authored pull requests** from December 2024 to April 2025. Specifically, the number of AI-authored PRs increased from 550 in December 2024 to 2,426 in January 2025, followed by continued growth in subsequent months (3,273 in February, 3,775 in March, and 5,428 in April).

This steady increase indicates a rapid adoption of AI coding agents in software development workflows over a short period of time.

	year_month	num_ai_prs
0	2024-12	550
1	2025-01	2426
2	2025-02	3273
3	2025-03	3775
4	2025-04	5428

```
np.float64(56524.24999999999)
```

The **positive coefficient** indicates a strong upward trend, meaning that the number of **AI-authored pull requests increases** substantially over successive time periods.

This confirms sustained growth in AI-generated contributions over time.

GITHUB LINK: https://github.com/mounatounakti/Acceptance-and-Trends-of-AI-Authored-Pull-Requests-on-GitHub