542 AIDevSet Project

YUE WANG*, University of British Columbia at Okanagan, Canada
YUSEN RONG†, University of British Columbia at Okanagan, Canada

ACM Reference Format:
Yue Wang and Yusen Rong. 2025. 542 AIDevSet Project. 1, 1 (December 2025), 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

With the rise of artificial intelligence (AI) in recent years, there has been a lot of discussion about if and how soon AI can replace software engineers in most companies. Historically, the code generated by AI can be questionable, especially if the language is relatively new or not as popular. Several analysis will need to be done on the capabilities of AI agents before even broader adoption by companies, and is probably a yearly exercise for every company to evaluate effectiveness and performance. In this report, we will use some of the AIDev datasets to analyze some performance and adoption metrics to see where the state of AI agents is currently. The GitHub project for this report is located at https://github.com/yueywangse/542Project.

2 Analysis of AI Agent Usage VS Coding Language

The first question that comes to mind is which coding language uses AI agent Pull Requests more often? Is there a specific tendency to adopt AI agents more in some languages vs others? This data is very interesting and might give us some insight on what language AI agents work best, where they need improvement, and if companies that utilize a certain language should start looking at adopting AI agent usage if they have not already.

2.1 Methodology

To determine the frequency of AI agent usage per language, I will mainly use two of the AIDev datasets, pull_requests and repository. I first checked that there are no Na values inside each of the repositories so that I can combine the two based on the repo_id column, unfortunately we found that around 37 out of 2807 repositories do not have data on what language they use, and with such a low number affected and no reliable way to collect that data ourselves, I simply discarded those rows of data. After the merge I confirmed that 183 of the 33596 rows of pull request data was also lost, which is a small number as well and was acceptable to me. With the cleaned data, we can then calculate and plot the frequencies of repos per language, the frequencies of prs per language, and

*General Report structure, wrote Introduction and section 2 and subsections, wrote code for section 2 and subsection, proof reader
†

Authors' Contact Information: Yue Wang, yueywang@student.ubc.ca, University of British Columbia at Okanagan, Kelowna, British Columbia, Canada; Yusen Rong, , University of British Columbia at Okanagan, Kelowna, British Columbia, Canada.

ACM XXXX-XXXX/2025/12-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

finally divide the two statistics to get average number of prs per repo per language to estimate language adoption.

2.2 Results

The generated frequency graphs are as follows.

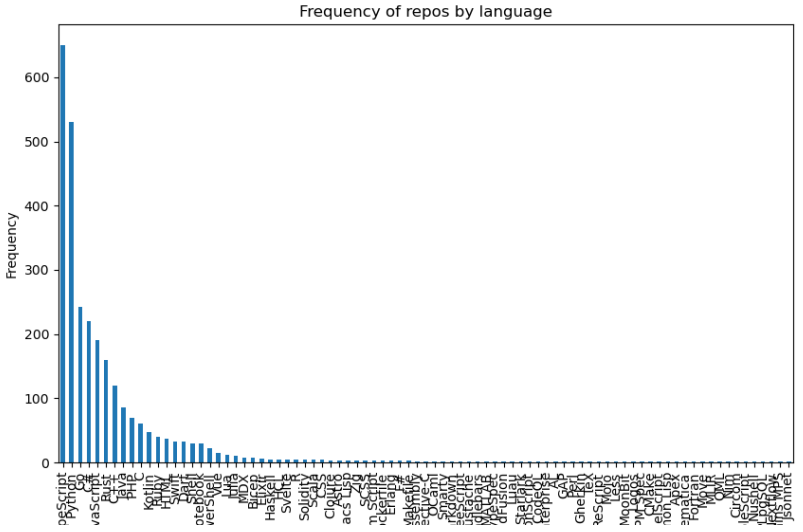

Fig. 1. Frequency Graph Depicting Number of Repos by Language

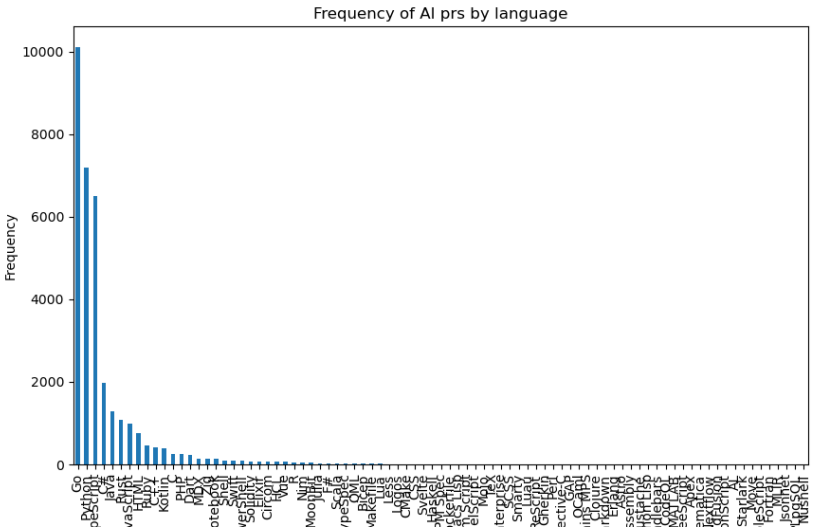

Fig. 2. Frequency Graph Depicting Number of AI Agent Prs by Language

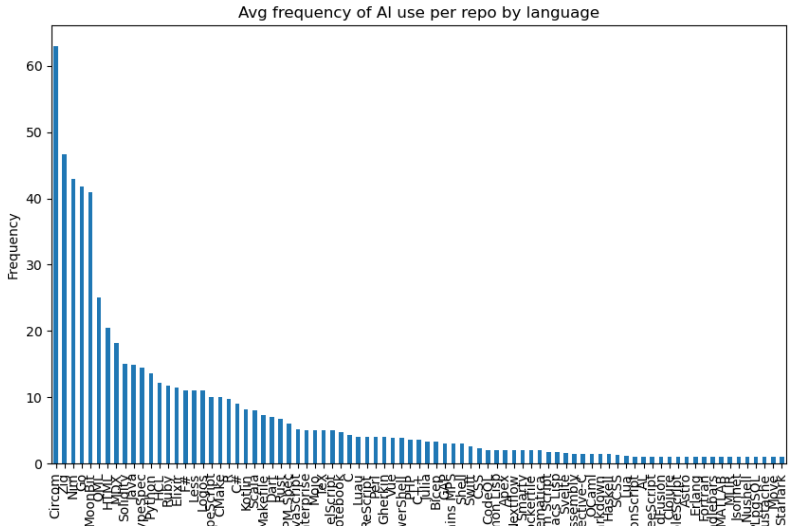

Fig. 3. Frequency Graph Depicting Average Number of AI Agent Prs per Repo by Language

2.3 Analysis

Looking at Figs. 1 and 2 above, we can see that the most popular languages to use in code repositories do not match the most popular languages to use ai agent prs on, which could have several causes. One of my guesses is that some languages are easier for ai agents to work with and thus have better models and be more developer friendly, while others are the opposite. Another guess is that companies that use those languages are more optimistic about AI usage and have pushed their employees to integrate more with AI agents than others.

The final Fig. 3 above shows the average amount of AI agent pull requests per repo separated by language. This can serve as an estimate of AI agent adoption per language for consideration, and might help a company decide whether they should follow suit/continue on their path of AI integration if they also use the same language. Do note however that this graph can be misleading, as the language with the largest adoption of AI agents, Circom, is used in an extremely small number of repos, so it might just be the effect one one person who is really into AI adoption. I would recommend only consulting the results for the more popular programming languages and not all of them.

3

4