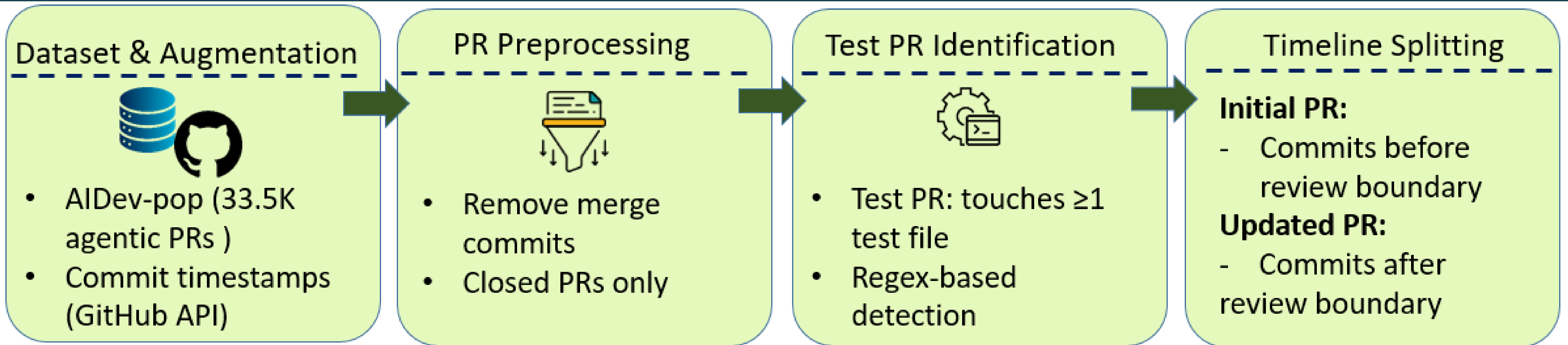


WHY TESTING MATTERS IN AGENTIC PRS?

- Autonomous coding agents increasingly submit pull requests.
- Testing is critical for correctness and maintainability
- It is unclear how tests appear in real agent-driven PR workflows

We analyze 33.5K agentic pull requests to understand when tests appear, how they evolve, and how test PRs differ from non-test PRs.

DATASET AND METHODOLOGY



RQ1: Test inclusion is increasing in agentic PRs

Test Inclusion Rate Over Time (2025)

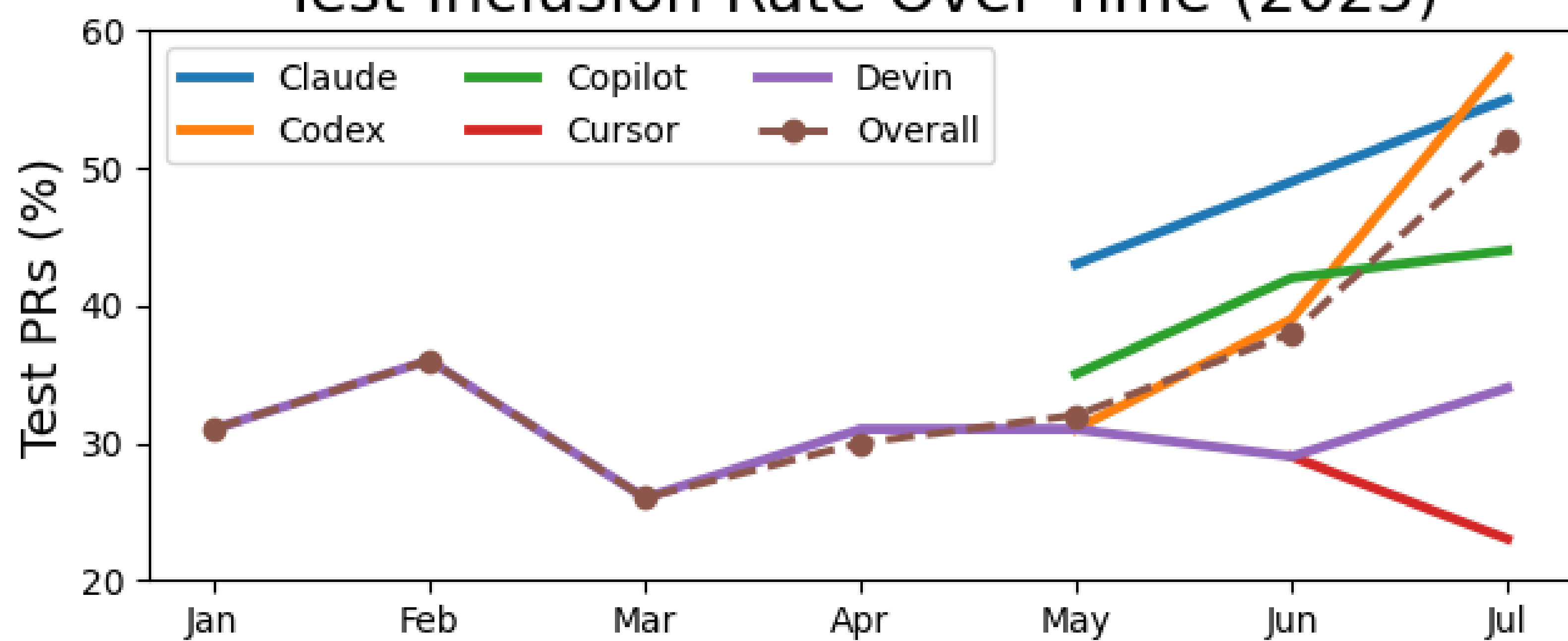


Fig.: Only months with ≥ 20 PRs per agent are shown.

Overall test inclusion rises from **31% to 52%** over time

Task Distribution Across Agents

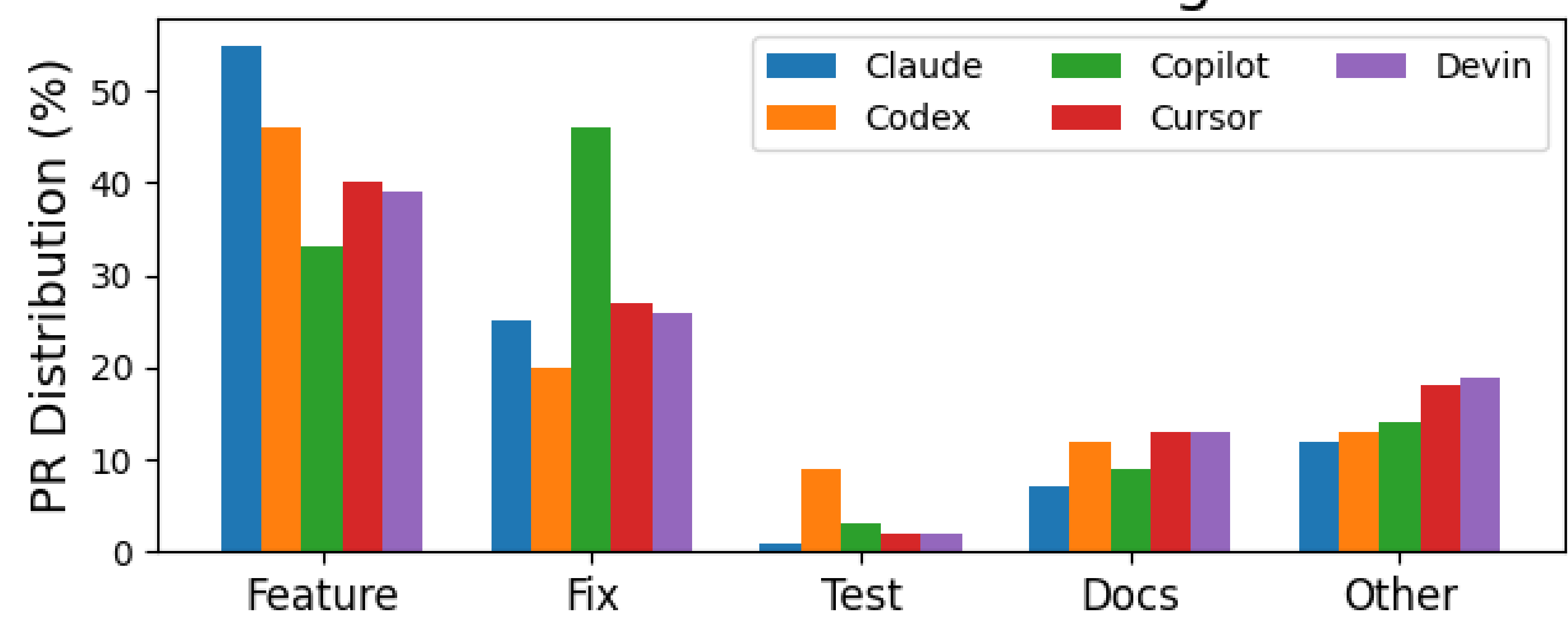


Fig.: PR type distribution.

Test inclusion aligns with task distribution: Feature-heavy (Claude, Codex) vs docs-heavy (Cursor, Devin)

RQ3: Test PRs are larger, have longer turnaround time

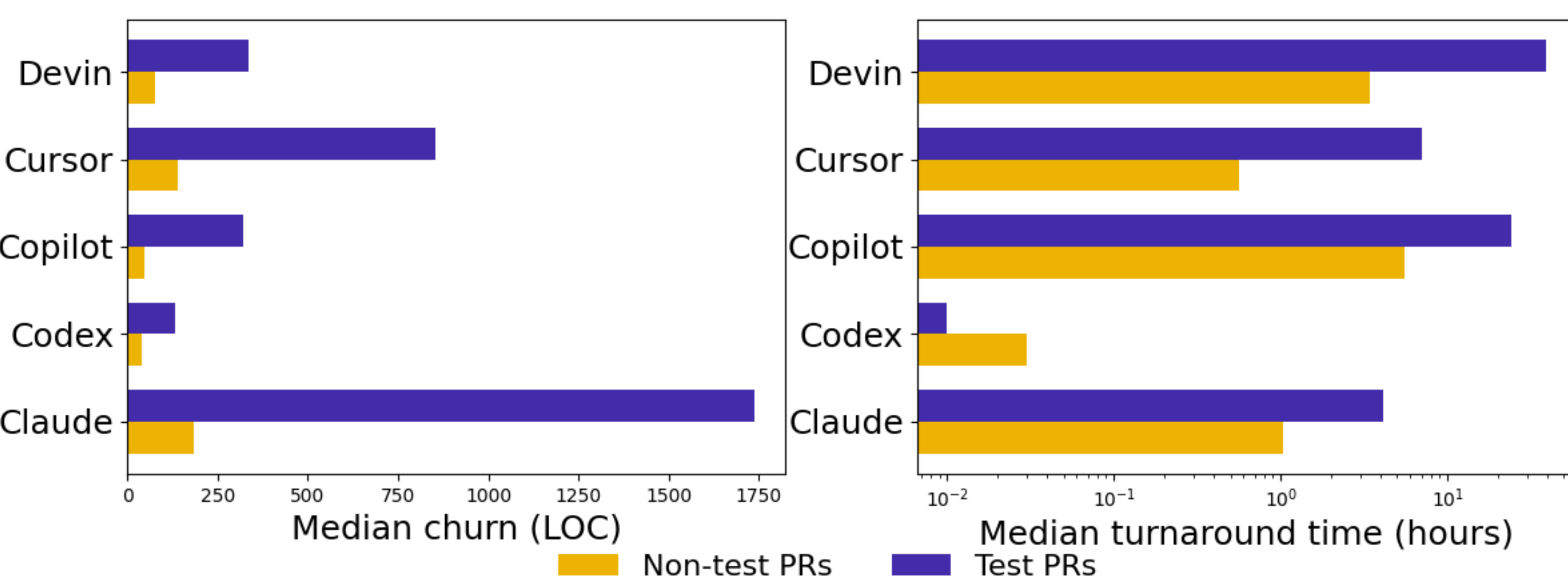


Fig.: Comparison for all Test and Non-test PRs

For PRs linked to GitHub issues, fraction of PRs whose linked issue was also closed (IC): indicate a stronger signal of human acceptance

Agent	Merge NT	Merge T
Claude	68.3%	74.3%
Codex	93.4%	81.5%
Copilot	64.4%	60.5%
Cursor	77.8%	86.2%
Devin	61.2%	26.4%

Table.: IC and Merge Rates for Issue-Linked PRs.

- Merge rates are largely similar across test and non-test PRs
- Exception: Devin less frequent test PR merge

RQ2: Tests often added early, frequently revised later

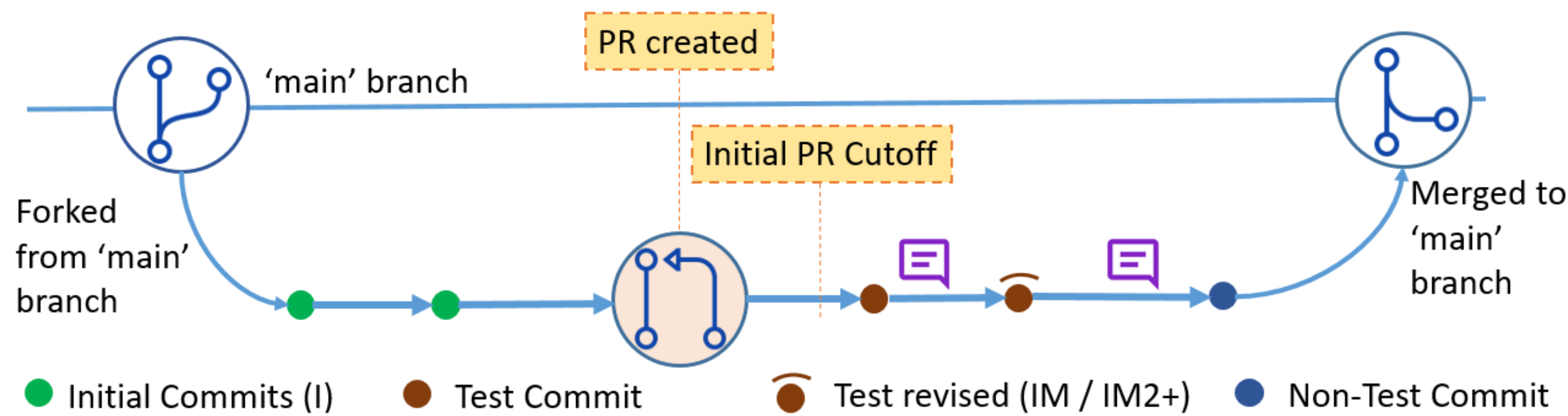


Fig.: An example PR timeline

Agent	Initial Only (I)	Later Only (L)	Both (I+L)	Revised (IM)	Multi-Rev (IM2+)
Claude	65%	14%	19%	34%	16%
Codex	96%	2%	2%	4%	1%
Copilot	70%	11%	18%	45%	24%
Cursor	64%	15%	19%	39%	16%
Devin	58%	24%	15%	59%	32%

Table: Test Introduction and Revision Timing in PRs

- Codex: mostly initial-only tests (96%)
- Devin: frequent later test revisions (59%)

LIMITATIONS

- Test PRs identified using filename and path heuristics
- The study does not evaluate test quality, adequacy, or coverage
- Descriptive findings, based on PR- and commit-level signals

REFERENCES

[1] Li, Hao, Haoxiang Zhang, and Ahmed E. Hassan. "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).

