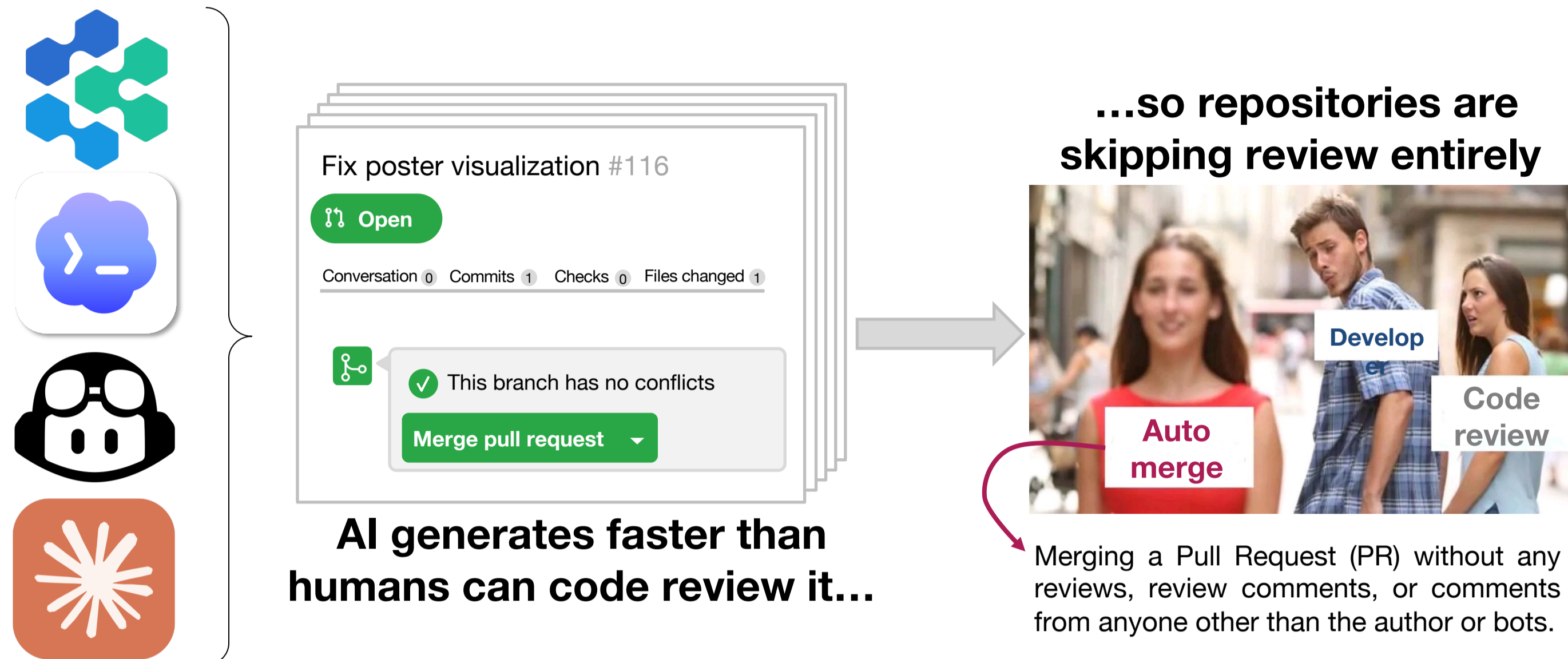


LGTM! Characteristics of Auto-Merged LLM-based Agentic Pull-Requests

Ruben Branco^{*†}, Paulo Canelas^{*‡}, Catarina Gamboa^{*‡}, Alcides Fonseca[†]

[†] LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal
[‡] Software and Societal Systems Department, Carnegie Mellon University, US
^{*} Authors contributed equally to this work



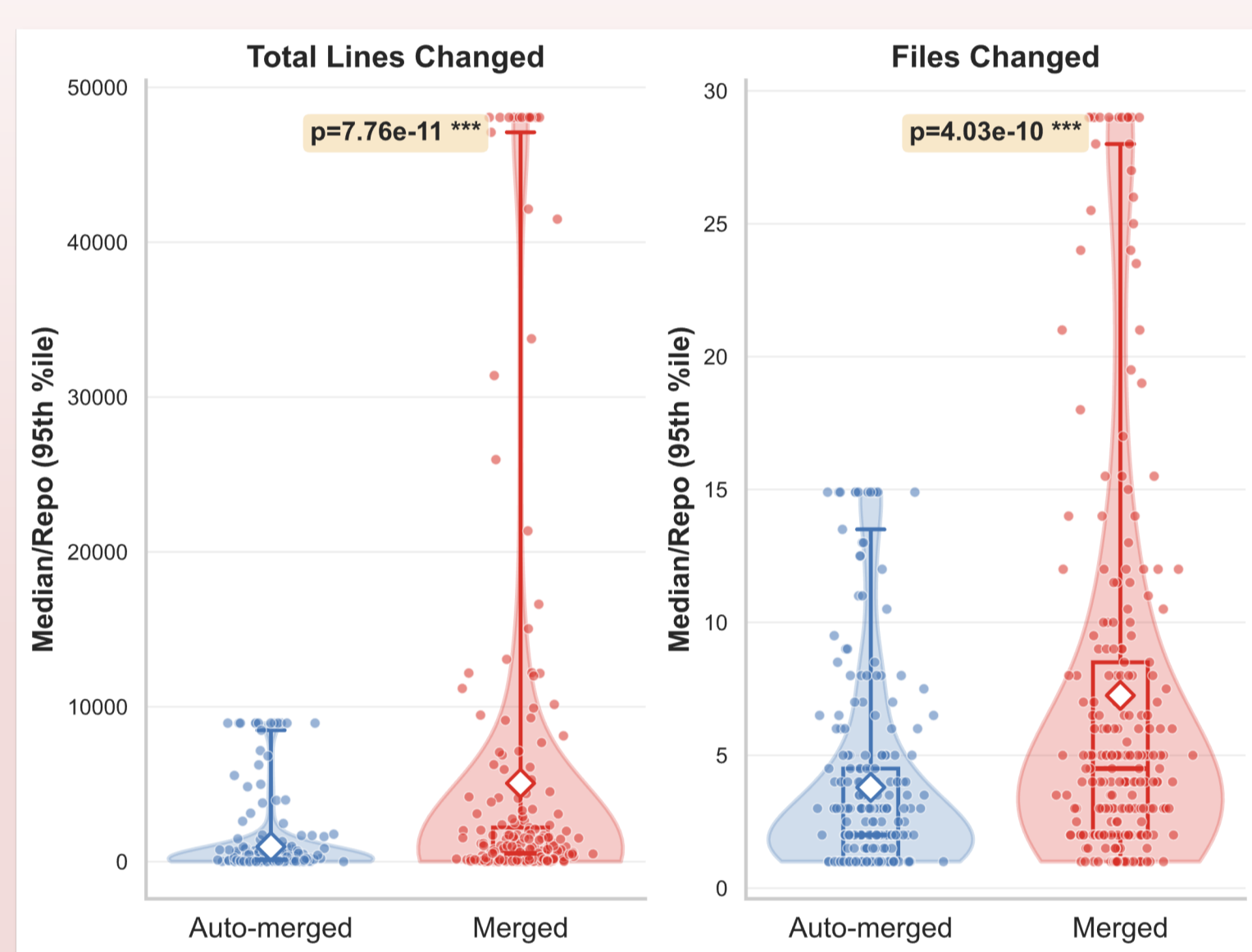
RQ1. What properties of agentic pull requests and repositories relate to auto-merges?

RQ2. What characteristics distinguish human and agent authored auto-merged PR?

AIDev dataset **5** agents **33,000** pull requests **2,800** repositories **≥100** stars

Answer to RQ1.

Auto-merge is bimodal: repos either do it for all agentic PRs or none. Auto-merged PRs are smaller, in less mature repos, and for lower-risk tasks.

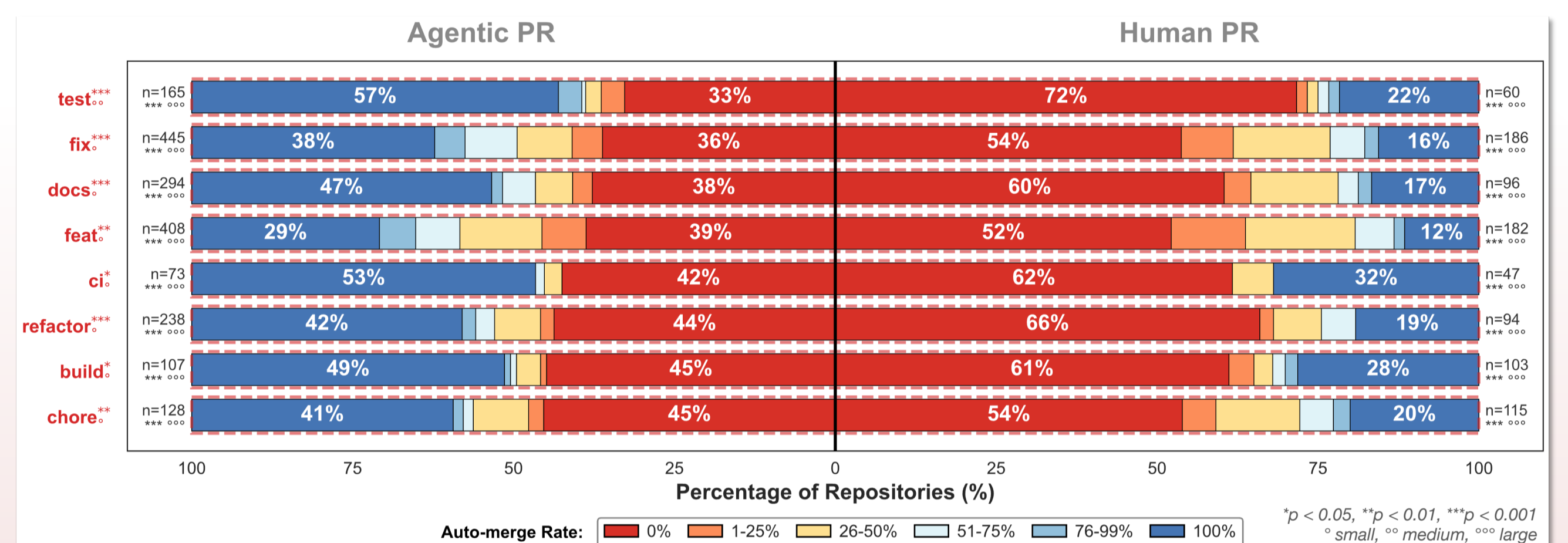


Agentic auto-merge follows a **bimodal distribution**, repositories either merge or avoid it.

Auto-merge PRs are **significantly smaller** across all size metrics: lines, files, additions, and deletions.

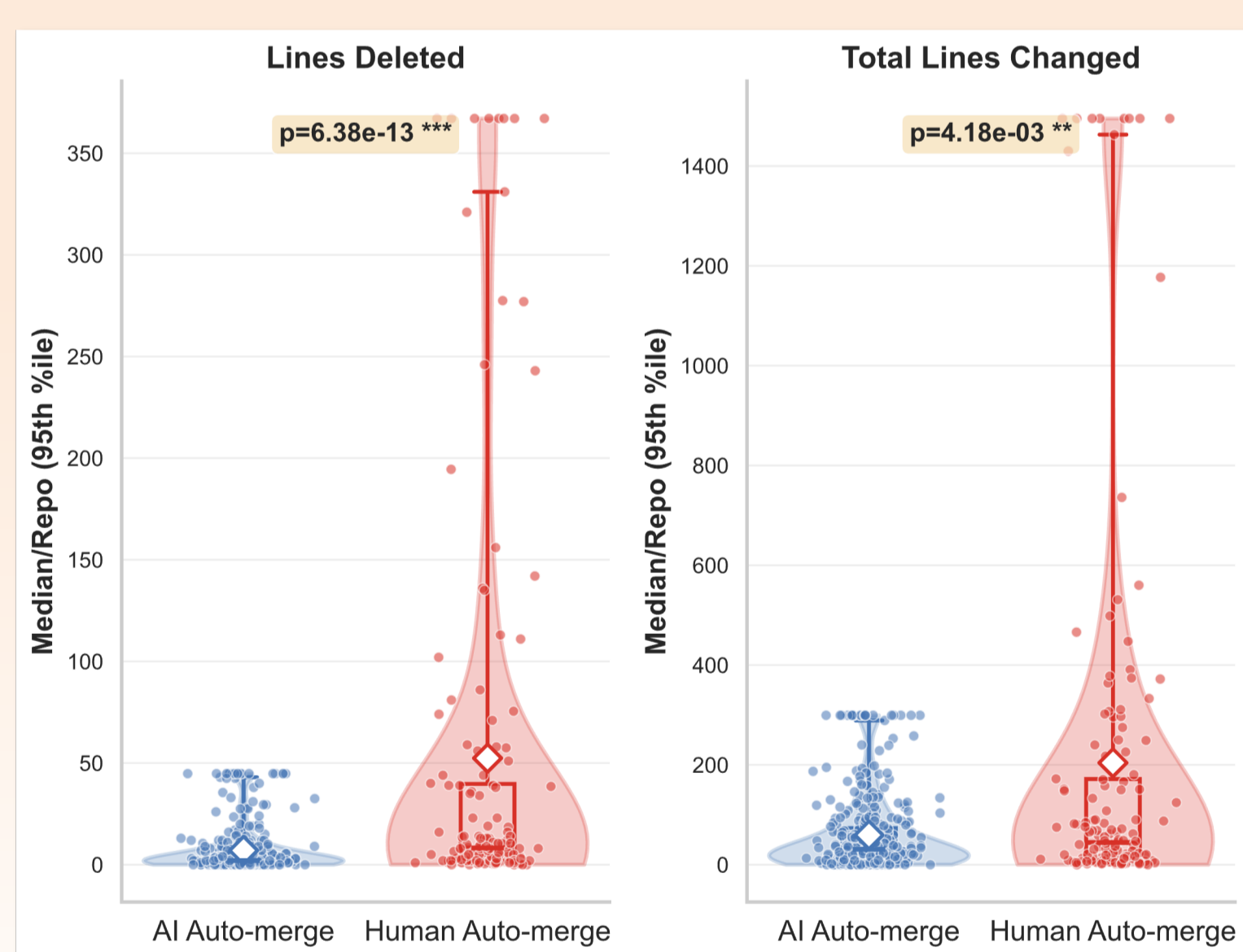
Tests are highly trusted, features are less: tests and CI reach 100% auto-merge in over half of repos, while features reach only 29%, likely reflecting their higher risk.

For all task types, more repositories reach 100% auto-merge for agentic PRs than for human PRs.



0% auto-merge rates occur in more mature repositories with more stars, forks, CI checks, and higher health scores.

Auto-merge is all-or-nothing for agentic PRs. When maintainers allow it, AI code skips review more often than human code, unless agents delete.



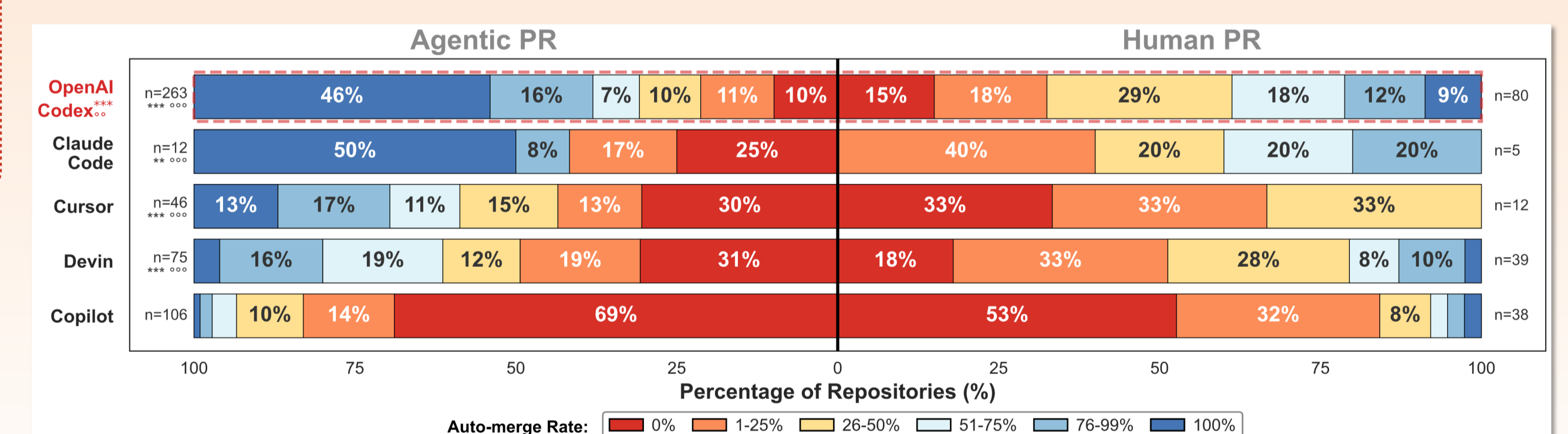
Agentic auto-merged PRs are **smaller and delete less** code than human ones.

Agentic PRs with deletions are auto-merged less often, suggesting **caution toward agent-removed code**.

In the same repositories, only OpenAI Codex shows significantly different auto-merge rates than human PRs.

Answer to RQ2.

Agent PRs are smaller and achieve higher auto-merge rates than human PRs. However, maintainers show caution toward agentic PRs that delete existing code.



Not all agents are the same: Codex and Claude have higher auto-merge rates. AI assistants and editors, Copilot, Cursor, and Devin have less.

Actionable Takeaways

1 **Keep agent PRs small and single-purpose**
Auto-merged agentic PRs are consistently smaller and more focused, so developers should scope AI-generated changes to one clear task, with minimal files and lines changed.

2 **Start with low-risk, developer-facing tasks**
Developer-facing tasks (tests, CI, builds) are the safest bet for agentic auto-merge. Customer-facing changes are riskier and should receive human review.

3 **Treat code deletion as a review trigger**
Maintainers are notably more cautious when agents remove existing code, so PRs with deletions may benefit from extra explanation, or mandatory human review.

Mining Challenge
#116

More details in the paper



<https://doi.org/10.1145/3793302.3793621>