

# Milestone 1: Research Questions and Methodology

Group Member: Chongwen Sun, Inaara Rajwani

## 1. Introduction

AI coding agents (e.g., Claude Code, Devin, Copilot, Cursor) now submit real pull requests (PRs) in GitHub repositories. Their large-scale behavior, reliability, and collaboration patterns are still poorly understood. This milestone defines three research questions (RQs) and the corresponding methodology using the AIDev dataset.

## 2. Research Questions

**RQ1: Testing Behavior** — How often do AI agents include tests in their PRs, and how does testing coverage differ across agents? *Motivation:* Test activity reflects reliability and is a critical indicator of software quality.

**RQ2: Review Dynamics** — What review comment categories do Agentic-PRs receive (correctness, style, documentation, testing), and which types are hardest for AI agents to address? *Motivation:* Reveals weaknesses of AI-generated patches and how humans respond to them.

**RQ3: Acceptance Prediction** — Can early PR features (description length, churn, files changed, agent type, presence of tests) predict acceptance vs. rejection? *Motivation:* Early signals help maintainers manage workload and guide future agent design.

## 3. Methodology

**Dataset.** We use the AIDev dataset (PRs, comments, reviews, commits, agent labels). Tables are joined via `pr_id` and repository metadata.

**RQ1.** Detect test files (e.g., `test/`, `tests/`, `_test.py`). Compute: (1) %PRs containing tests, (2) test-to-code churn ratio, (3) differences across agents.

**RQ2.** Categorize review comments via keyword rules (e.g., “bug”, “test”, “style”, “security”). Measure frequency and whether follow-up commits resolve comments.

**RQ3.** Extract features (such as churn, number of files changed, description length, presence of tests, etc.) and analyze whether these features show meaningful trends with respect to PR acceptance.

## 4. Expected Contributions

(1) Quantitative measurement of AI agents’ testing behavior. (2) Identification of comment types challenging for AI agents. (3) A simple early acceptance model to improve human–AI collaboration.