

# Evaluating AI Code Review Tools

Real-world bug detection across five AI review tools.

## TOP PERFORMER

### CodeRabbit

PRECISION

**95.88%**

High-confidence findings

CRITICAL BUGS

**Most Identified**

25 critical bugs

CONSISTENCY

**5 of 6 Repos**

Led in critical bug detection

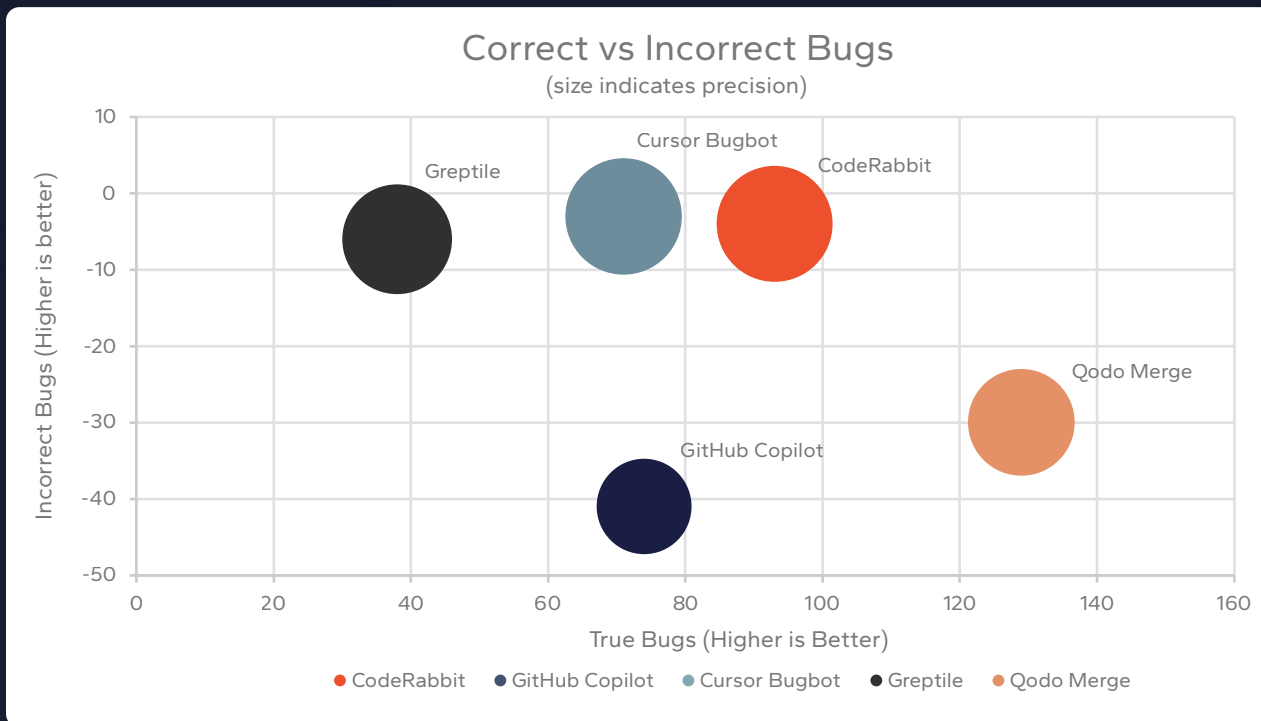
## WHY IT MATTERS

The most effective code review tools maximize identification of high-impact bugs while minimizing noise.

## Signal vs Noise

More findings ≠ better code reviews.

The strongest tools maximize real bugs found while minimizing false positives.



CodeRabbit pairs high true bug detection with very low false positives relative to competitive tools.

## HOW WE TESTED

5 tools • 6 repos

60 real bugs (historical PRs)

Manual grading

## WHY CODERABBIT

25 critical bugs (most)

95.88% precision

Led in 5/6 repos

## Precision vs Critical Impact

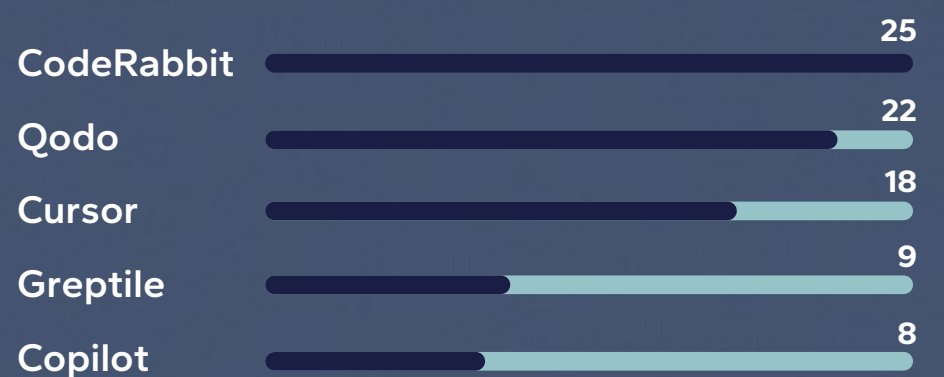
### PRECISION

Share of findings that are real issues. Higher precision means less time wasted on false positives.



### CRITICAL BUGS

High-impact issues like crashes, security risks, and incorrect results – the bugs that matter most.



## Key Takeaway

Accuracy + impact beats raw volume.

CodeRabbit delivers the strongest overall balance of precision, critical bug detection, and consistency across repositories – making it the clearest “high signal, low noise” option in this study.