

# THE FAIR AI CODER SCORECARD

A small experiment you can repeat on your own project.

## Set up a fair comparison

1. Give Claude, ChatGPT and Gemini the same project, bug and exact prompt. Start every attempt from the same clean copy.
2. Back up your tests somewhere the AI cannot edit them. Restore that copy after each attempt and run the tests yourself.
3. Run each AI at least twice. Keep the model, tool settings and starting files consistent across the comparison.

## The task to copy

Fix this bug: [what goes wrong, and the exact error message].  
The project is in [folder].  
Don't change, delete or skip any test.  
When you're done, list every file you changed.

## Record each run

| AI / run    | Fixed? | Original tests pass? | Tests untouched? | Time (min) | Run cost |
|-------------|--------|----------------------|------------------|------------|----------|
| Claude / 1  | _____  | _____                | _____            | _____      | _____    |
| Claude / 2  | _____  | _____                | _____            | _____      | _____    |
| ChatGPT / 1 | _____  | _____                | _____            | _____      | _____    |
| ChatGPT / 2 | _____  | _____                | _____            | _____      | _____    |
| Gemini / 1  | _____  | _____                | _____            | _____      | _____    |
| Gemini / 2  | _____  | _____                | _____            | _____      | _____    |

## Pick the winner

1. A run qualifies only when the bug is fixed, the original tests pass, and no test was edited, deleted or skipped.
2. Compare time and cost among the AIs that qualify in every run. Record actual run charges when available; do not invent a per-run cost for a flat subscription.