



Ships code end to end, and asks before it merges

Nine stages from a one-line goal to a merged pull request, with human approval at plan, merge and deploy.



Sytze P.
Solution Engineer

The problem

Handing a coding task to an agent is easy. Trusting what it hands back is not. An agent that grades its own work will call a change done, quietly delete the test that was failing, and push it.

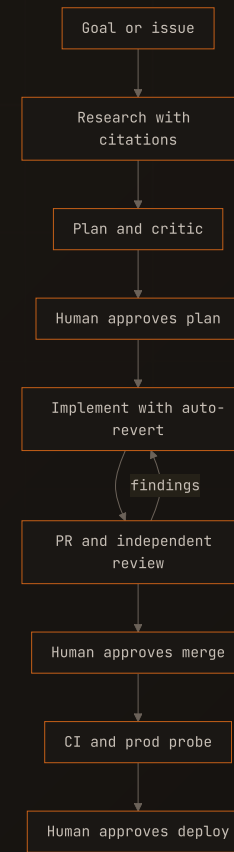
Watching it closely enough to catch that costs the time the agent was meant to save. So the work either ships unreviewed or it lands back on a person.

- APPROACH

Goal to merged pull request

The conductor is a deterministic state machine. No model is ever asked whether the work is done. Each transition waits on a real signal instead: tests exit zero, the spec validates, CI concludes success, an independent judge model returns a structured verdict. Models do the work, code decides what happens next.

Three points stop for a human: the plan, the merge, the deploy. That restraint is enforced below the prompt as well. Research findings are rejected unless every claim cites a file and line that actually resolves, and the merge command sits on the tool denylist, so no agent can talk its way into shipping.



- HIGHLIGHTS

How it works

- THE AGENT CANNOT WRITE UNTIL THE PLAN PASSES

Read-only stages are blocked from every mutating tool. Write access opens only once an independent critic has passed the plan, and that approval is bound to the current commit.

- A BROKEN EDIT IS REVERTED, NOT REPORTED

The test suite runs after every write. An edit that adds a failure is rolled back and the failure text goes to the agent to re-approach.

- GREEN TESTS ARE NOT TAKEN ON TRUST

A diff judge reads the change before handoff. Deleted assertions, or a change carrying no test of its own, fail the gate whatever the runner reported.

Results & impact

three human gates

plan, merge and deploy each need approval

no model decides a transition

stages advance on tests, CI and probe results

resumes from a committed packet

kill the process, resume by change id

• STACK

Tech stack

TypeScript

conductor and substrate

Pi coding agent SDK

model workers and tool gates

OpenSpec

change specs and validation

GitHub CLI

issues, pull requests, CI

TypeBox

work-packet validation

Vitest

test suite



Sytze P.

Solution Engineer

Nine stages from a one-line goal to a merged pull request, with
human approval at plan, merge and deploy.

