StephOS

AI PRODUCT & SYSTEMS DESIGN · CASE STUDY 03

HOW I SHIP WITH AI AGENTS. I decide.

A designer taking ideas to installed, working software, with one AI building and another checking.

Steph Campagnoli
Product direction, specifications, acceptance criteria
StephOS · August–October 2026



01 / THE BRIEF

I DON'T WRITE THE CODE. I *decide* WHAT SHIPS.

AI coding agents can produce a lot of software quickly. The hard part is knowing whether it does what you meant, and whether it broke something you already had. As a designer without an engineering background, I needed a way of working where "done" didn't depend on reading code or trusting a confident summary.

THE PERSON

A designer directing AI agents on a real, long-running product.

THE TASK

Take each change from an idea to an installed release, without losing data or breaking what already worked.

THE CHALLENGE

Make quality something I can see and check, not something I have to take on faith.

02 / WRITE THE BEHAVIOR FIRST

EVERY CHANGE STARTS AS A PROPOSAL I *approve.*

Each change is written up as one small slice: what it does, what it deliberately doesn't do, and how we'll know it works. I approve it in my own words, often with conditions.

A FINDING THAT CHANGED THE DESIGN

A design once treated half-finished reply fragments as "progress." I rejected it: a fragment isn't an update. It became a real lookup step that has to report what it found, to me, in plain words.

1 **PROPOSAL**
One slice, scoped, approved in words.

2 **DESIGN REVIEW**
Checked before any code exists.

3 **BUILD**
Claude Code builds it, with tests.

4 **REVIEW**
Codex reviews the exact change.

5 **INSTALL**
Ten checked steps, with a way back.

6 **PROVEN**
Seen working in real use, or marked "not yet."

03 / TWO AIS, TWO JOBS

ONE BUILDS. A DIFFERENT ONE CHECKS.

5 of 5

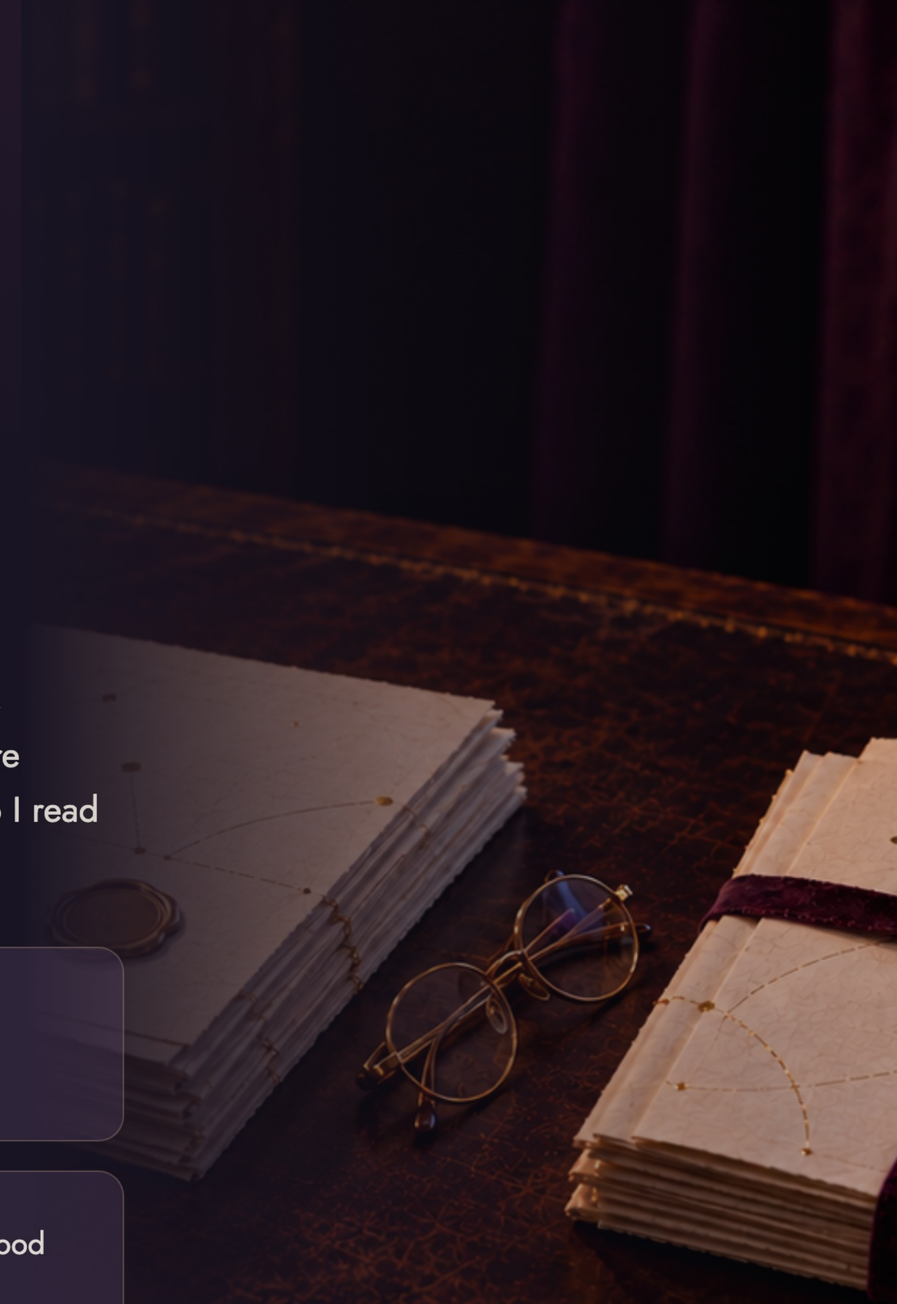
In each of the last five releases, the reviewer's first verdict was "required changes": 15 problems in all, every one fixed before install (as of October 24, 2026). I relay the reviews myself, so I read every finding.

"Stop sharing with Cloud" didn't stop automatic sends in chats approved the old way.

Damaged approval data could still let a message send.

Hiding the assistant's bookkeeping could also hide part of my own message.

A failure reading progress notes could turn a good answer into a failed one.



04 / MAKE “DONE” OBSERVABLE

TEN CHECKED STEPS FROM APPROVED TO INSTALLED.

Approval isn’t installation. Every release goes through the same ten steps, and I unlock the app myself at the install step.

REQUIREMENT	HOW IT’S CHECKED
EXACTLY WHAT WAS REVIEWED	The installed files match the reviewed change, byte for byte.
NOTHING ELSE CHANGED	Data, memory and permissions are compared before and after: identical.
NOTHING LOST	File counts match (2,240 before, 2,240 after); only the expected files were rewritten.
A WAY BACK	A recovery copy of the previous version is kept for every install.
MY *approval*	The install can’t finish until I unlock the app.

05 / INSTALLED ISN'T PROVEN

I TRACK INSTALLED AND PROVEN SEPARATELY.

A feature can be installed and still not be shown working in real use. Several of mine honestly say "not yet" or "synthetic check only," and one says "unverified."

CAPABILITY	INSTALLED · DEMONSTRATED IN USE
AUTOMATIC RECALL	Installed · Yes, in my own use
PROGRESS NOTES	Installed Oct 19 · First live use; usefulness still my judgment
STANDING CLOUD APPROVAL	Installed · Not yet used



A FAILURE, AND THE RULE IT LEFT

I once made a results file read-only to protect it; the next run couldn't write to it, and it looked like a provider problem. Now finished files are frozen by recording a fingerprint of their contents, never by locking them.

06 / WHAT IT ADDS UP TO

A WAY OF WORKING, NOT A ONE-OFF.

240 changes since August 14

29 installed versions kept, each restorable

39 dated decision records

Counts as of October 24, 2026

> AI BUILDS IT. AI CHECKS IT.
> I *accept* IT.

PROVEN IN MY OWN USE

- Proposals approved in words, with conditions
- Independent review on every release
- Ten-step installs with recovery copies

NOT YET

- No external users or team yet
- The reviewer is an AI, not a human engineer
- My numbers aren't compared with anyone else's

SOFTWARE YOU CAN *trust.*

SPECIFY BEFORE BUILDING

Behavior, boundaries and checks written down before any code exists.

DIRECT AI, VERIFY ITS WORK

Agents build and review; I accept only what I can open and check.

SAY WHAT ISN'T PROVEN

Installed and demonstrated kept apart, with limits stated plainly.

DISCUSS A ROLE OR COLLABORATION

hello@stephos.studio · stephos.studio · linkedin.com/in/steph-campagnoli

Steph Campagnoli