AI PRODUCT & SYSTEMS DESIGN · CASE STUDY 01

CONTINUITY WITH *consent.*

You decide.

Designing AI memory people can permit, inspect and correct.

Steph Campagnoli
Product direction, interaction design, requirements
StephOS · January–October 2026

01 / THE PROBLEM

AI KEPT SAYING IT REMEMBERED.

Every new chat started from zero. And when an assistant did "remember," it sometimes invented details, claimed work was saved when it wasn't, or replaced my original words with its own summary. For long-running work, that's worse than forgetting: you can no longer tell what to trust.

FORGETTING IS AN INCONVENIENCE. FALSE MEMORY IS A PROBLEM OF *trust*.

SO I SET THREE FOUNDING REQUIREMENTS

ORIGINAL WORDS ARE EVIDENCE

Summaries and decisions link back to what was actually said.

APPROVAL IS EXPLICIT

Nothing becomes approved on its own. Discussion is not permission.

DONE MEANS I CAN OPEN IT

A receipt or a claim is not proof. The saved record has to open and check out.

02 / MY ROLE

I DESIGN THE BEHAVIOR. AI WRITES THE CODE.

Sole designer and tester, January–October 2026. I set the goals, requirements and interaction direction, compared models, and decided what met acceptance. AI wrote the code and assisted research and documentation.

1

DEFINE

Name the failure people work around and what it costs them.

2

SPECIFY

Write the behavior, the boundaries and how failure should look.

3

DIRECT

Guide AI coding agents through implementation.

4

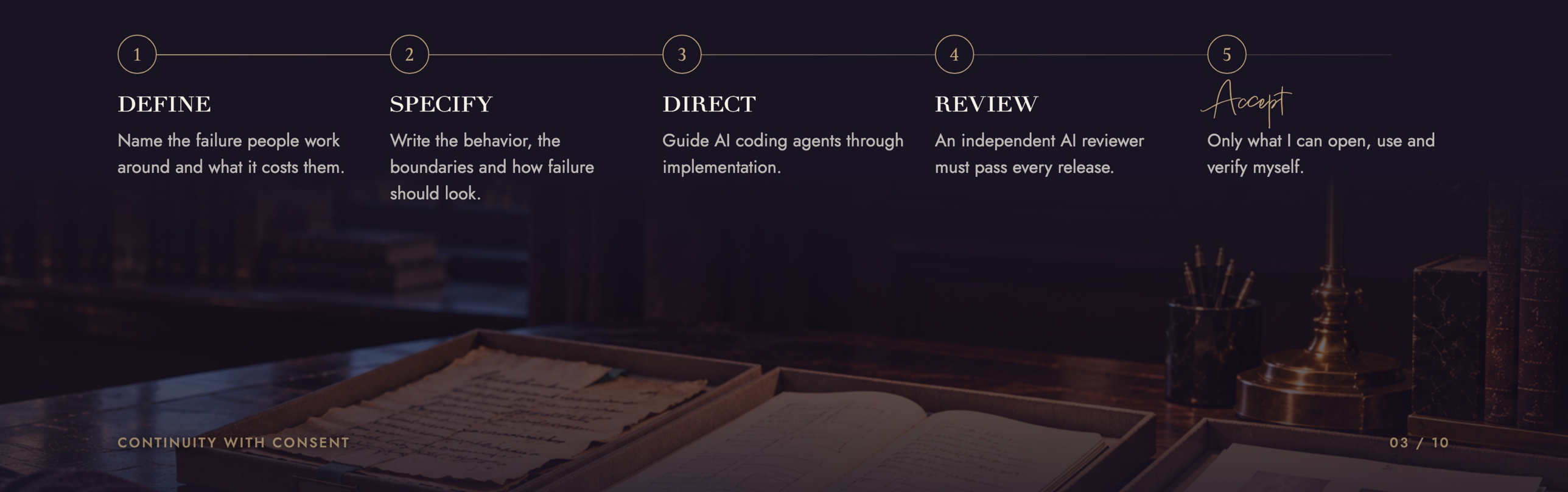
REVIEW

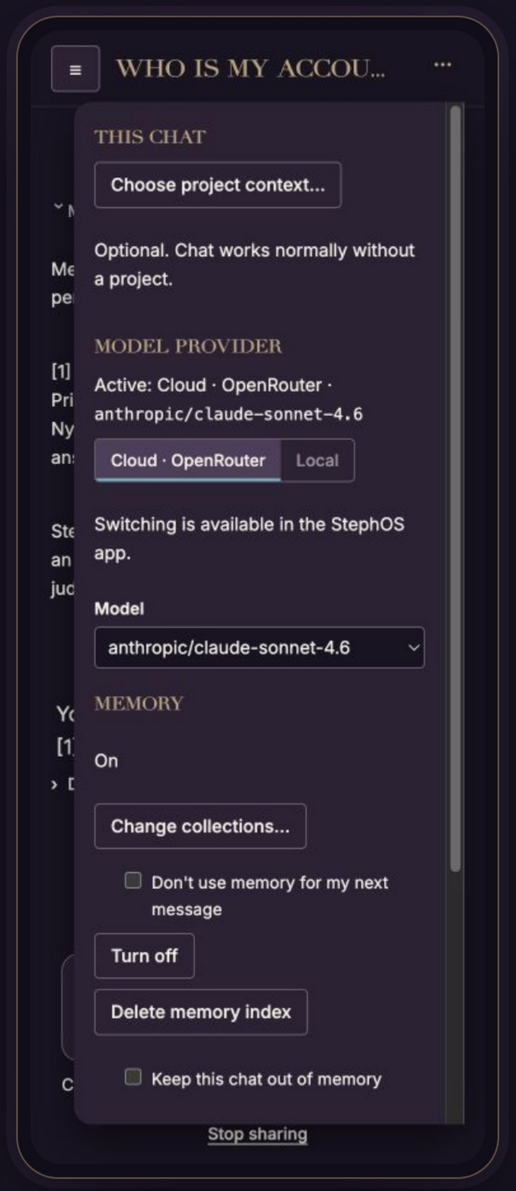
An independent AI reviewer must pass every release.

5

Accept

Only what I can open, use and verify myself.





Interface capture · synthetic records · scripted replies

03 / INTERACTION

ONE HOME FOR MEMORY *decisions.*

SEPARATE PERMISSIONS

Searching on this Mac and sending to a cloud model are different grants. Allowing one never implies the other.

DISTINCT CHOICES, NOT ONE SWITCH

Turn memory off, keep a chat out, or skip just the next message.

REACHABLE, NOT REPEATED

Controls live in one place instead of under every message.

04 / PROGRESSIVE DISCLOSURE

CONVERSATION FIRST. EVIDENCE ON REQUEST.

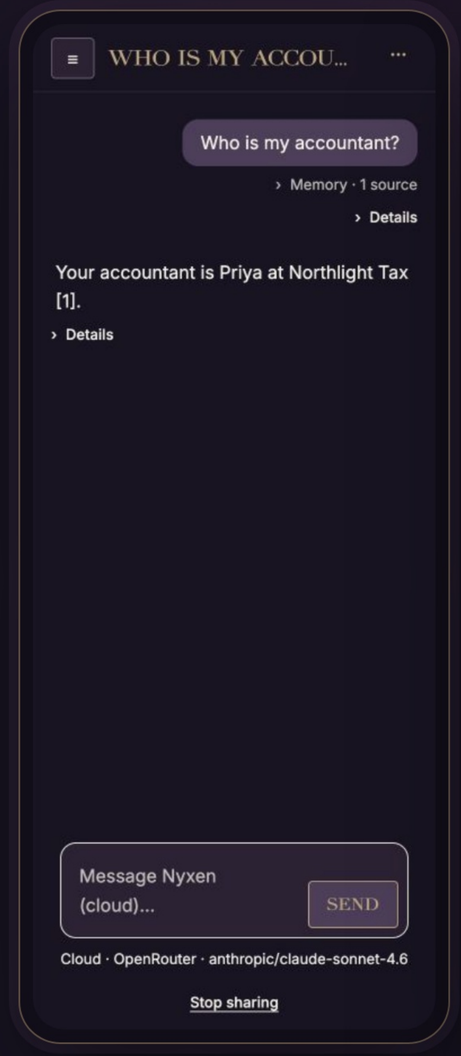
At rest, one quiet line: “Memory · 1 source.” Open it and you see the exact excerpt, who said it and when, and how the assistant used it.

THE INTERFACE STATES ITS OWN LIMIT

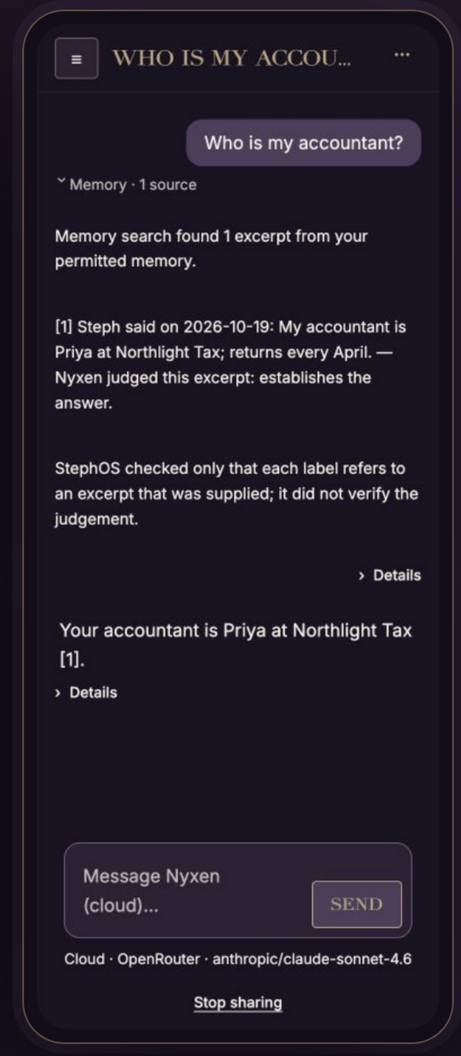
“StephOS checked only that each label refers to an excerpt that was supplied; it did not verify the judgement.”

HONESTY IS A DESIGN SURFACE, NOT A FOOTNOTE.

AT REST



ON REQUEST



Interface capture · synthetic records · scripted replies

05 / SYSTEMS

RETRIEVAL ONLY WITH *permission.*

Every answer that draws on memory passes the same path. Each step is a boundary the person can rely on.

1

PERMITTED COLLECTIONS

Only what the person allowed, for this use, can be searched.

2

SOURCES KEEP IDENTITY

Each excerpt stays tied to the chat, note or saved work it came from.

3

HISTORY STAYS GROUPED

A replacement never erases what it replaced.

4

CHECKED AT SEND TIME

Permission and source are rechecked just before anything leaves.

5

ANSWER ≠ APPROVAL

Review and Save stay separate, deliberate acts.

TRADE-OFF Automatic recall removes repeated manual choices but adds consent, relevance and index upkeep. I chose to carry that complexity in the system, not hand it to the person.

06 / EVALUATION

RIGHT ANSWERS. UNINVITED MEMORY.

18/18

facts recalled correctly

2 of 4

unrelated questions still drew on memory

MEMORY YOU DIDN'T NEED IS STILL DISCLOSURE. RELEVANCE IS A *consent* PROBLEM.

WHAT I CHANGED

The assistant now reads the recent conversation to resolve references, topic changes and corrections, searches with a specific subject, and asks when something is unclear. Sources reused from earlier answers are rechecked and capped. Shipped October 19; now being judged in everyday use.

24 synthetic cases across two versions. Not a claim of 100% accuracy.

07 / QUALITY

NOTHING SHIPS ON ITS OWN WORD.

My rule: no release installs without a passing independent review. The conversation release needed one fix before it passed. The next feature, durable progress notes, went through three reviews and five fixes before it was confirmed and installed.

PROGRESS NOTES · CONFIRMED, INSTALLED OCT 19

A DEFECT WORTH LEARNING FROM

Withdrawing a later correction could make an older value look current again.

NEW RULE

Undo never silently restores. Bringing a value back takes an explicit new correction.

REVIEW 1

Three defects found · all fixed

REVIEW 2

One subtler defect · fixed

REVIEW 3

One more edge case · fixed

CONFIRMATION

Passed · installed Oct 19

08 / HONEST SCOPE

WHAT'S PROVEN. WHAT ISN'T YET.

PROVEN IN MY OWN USE

- Permitted automatic recall across chats
- Memory controls in one place
- Original words preserved and linked to answers
- Every release independently reviewed before install

NOT YET

- No external users yet
- Relevance is improved, not solved
- Imported chat history is out of scope
- The conversation release is still being judged in use

WHAT I'D TEST WITH REAL PEOPLE

I haven’t run this yet. The study I’d run: a few people resuming real work. Can they tell where an answer came from, choose what memory may use, and recover from an irrelevant recall? Success defined before the first session.

AI THAT *people* RELY ON.

TURN AI FAILURES INTO REQUIREMENTS

Find the problem people work around and specify what the system must do instead.

DESIGN CONSENT AND EVIDENCE IN

Make permission, sources and limits part of the interface, not the fine print.

SHIP THROUGH REVIEW

Independent checks before release, and plain words about what isn't proven.

DISCUSS A ROLE OR COLLABORATION

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

Steph Campagnoli