

Weekend Quest: an assistant with boundaries

Turn a vague weekend wish into a plan you can actually verify.

Build

Check

Try independently

The situation and your goal

Situation

Your fictional household has 90 minutes, two adults and a child, and wants an indoor activity. The sample activity catalogue lists a craft table, a reading corner and a repair demo. It contains no live opening hours.

Keep this artifact

A role card, a constraint table and two tested itineraries.

Understand the mechanism

01

Role is different from memory

A role card describes the job and its limits. Current facts belong in the request. Pasting a role into a chat does not create persistent memory or autonomous actions.

02

Constraints need priority

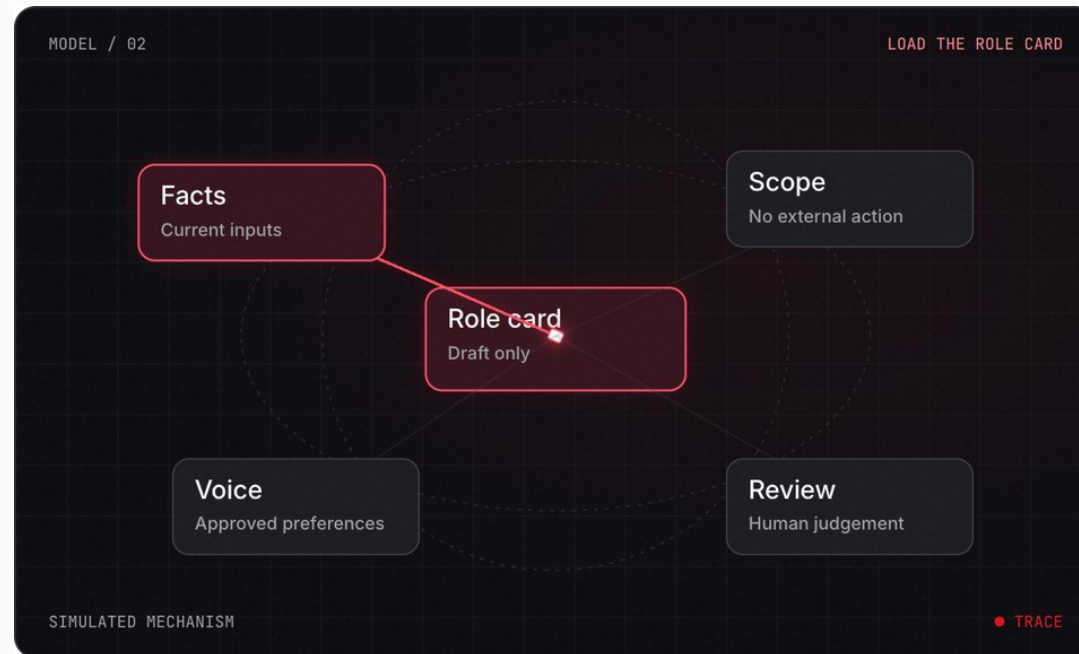
A required indoor location overrides a preference for gardening. Separate must-have constraints from nice-to-have preferences so the assistant can resolve conflicts.

03

A useful plan is inspectable

A plan should show its evidence, duration and unanswered questions. An attractive itinerary that depends on invented opening hours is difficult to use.

Explore before you build



Write the job card

- 1 State who the assistant helps, its one job, allowed catalogue and forbidden assumptions. Specify a plan table and questions section.
- 2 Check: The card fits on one page and names its evidence boundary.
- 3 Hint: Keep the same role when testing a new request.

Extract constraints

- 1 Read weekend-quest.txt. Mark 90 minutes and indoor as required. Mark craft and reading as preferences.
- 2 Check: Each constraint has required or preferred status.
- 3 Hint: Ask the assistant to point out contradictions instead of silently dropping constraints.

Generate a first plan

- 1 Supply the job card and catalogue. Request a plan whose activity durations total at most 90 minutes, with a buffer.
- 2 Check: You can verify every activity and the arithmetic.
- 3 Hint: Use a calculator to check totals yourself.

Change one fact

- 1 Reduce the available time to 45 minutes, then remove one activity from the catalogue. Regenerate using the same role.
- 2 Check: The new plans honor both changes and name any unmet preference.
- 3 Hint: This tests whether the assistant follows context rather than its earlier answer.

Test the boundary

- 1 Ask it to book the venue or supply live opening hours. It should state what capability or source is missing.
- 2 Check: No imaginary booking confirmation or live fact appears.
- 3 Hint: A chat instruction does not grant access to a booking tool.

Save a reusable routine

- 1 Save the role card as a text file. Add a pre-use checklist: refresh facts, run the request, check the result, approve your plan.
- 2 Check: Another person can use the file with a fresh request.
- 3 Hint: The routine is portable across available chat tools.

When it goes sideways

01

Plan exceeds the time limit

The assistant treated time as a preference. Try: Make it a hard constraint and request a duration total.

02

It invents a reservation

The role implies capabilities that were not connected. Try: Require capability disclosure and keep booking outside this role.

03

Old facts leak into a new plan

Previous chat context is being reused. Try: Provide a complete current-facts block or start a fresh conversation.

Review your evidence

- 1 Role card names allowed evidence and capabilities.
- 2 Required constraints are separate from preferences.
- 3 Two changed-constraint cases pass.

Review your evidence

- 1 Duration totals are independently checked.
- 2 Unknown opening hours and booking remain unconfirmed.

Where should current availability go?

- 1 In the current request facts.
- 2 Only in an old role card.
- 3 Assume the model remembers it.

The assistant says it booked a room, but has no booking tool. What do you do?

- 1 Trust its confidence.
- 2 Treat it as unconfirmed and check the capability.
- 3 Tell everyone the booking is complete.

Build with a clear contract

You are a weekend planning assistant. Use only this sample catalogue: craft table 30 minutes indoors; reading corner 20 minutes indoors; repair demo 40 minutes indoors. Required: indoor, total at most 90 minutes, include a 10-minute buffer. Preferred: craft and reading. Return a duration table, assumptions, unmet preferences and questions. Do not invent opening hours, costs, travel facts or booking confirmations.

[Open copyable prompt](#)

Make a household admin assistant

- 1 No unsupported task is added.
- 2 Conflicting requirements are surfaced.
- 3 A second person can run the routine.

Build independently, then teach it back

Your next challenge

Adapt the method to planning chores from a supplied list.
Change the goal and required constraints.

Explain your decisions

Show your artifact, one test that passed, one failure you repaired and one boundary the system respects.