

The setup planner: an agent you can inspect

Watch an agent choose a calculator, compare options and stop at its limit.

Build

Check

Try independently

The situation and your goal

Situation

Two fictional setup options cost $3 \times 18 + 2 \times 12$ and $4 \times 20 + 1 \times 15$. The approved budget is 90. The agent may use a Calculator tool; it has no shopping or weather tool.

Keep this artifact

An agent workflow, a tool-call trace and a three-case boundary report.

Understand the mechanism

01

Tool choice is the distinguishing step

The model can choose a connected tool and use its result. A fixed chain has a predetermined model call instead. Merely writing the word agent in a prompt does not connect capabilities.

02

Tool output is evidence

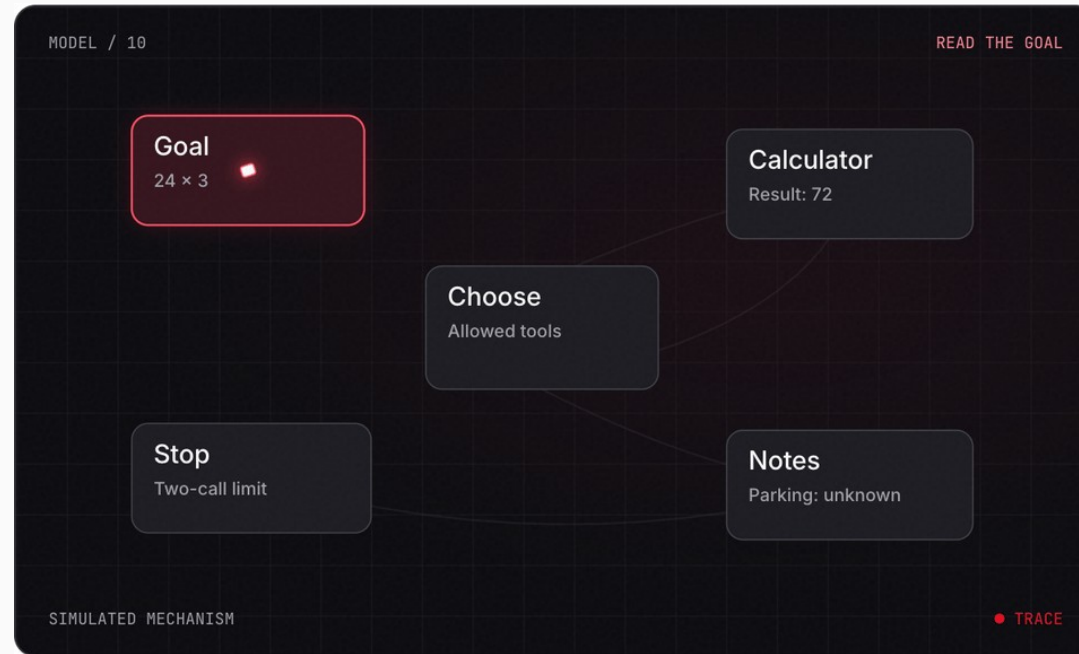
The arithmetic result should come from the calculator call and be visible in the trace. A plausible narrated tool call is not proof that an execution happened.

03

A stopping rule makes failures inspectable

Set a maximum iteration count and an error path. If the task needs an unavailable tool, the agent should explain the missing capability rather than inventing a result.

Explore before you build



Define tool authority

- 1 Write the task and whitelist: compare supplied costs, use Calculator, return a draft recommendation. It may not purchase, send or browse.
- 2 Check: The allowed actions and missing tools are explicit.
- 3 Hint: All prices are fictional sample inputs.

Connect the real agent

1

Configure n8n AI Agent with a model that supports tool calling and attach the Calculator sub-node. Set Maximum Iterations to 4 for this exercise.

2

Check: The node has at least one connected tool and a visible limit.

3

Hint: A chat model working in a basic chain may not support tool calling.

Run both options

- 1 Supply the two cost formulas and budget90. Inspect the calculator input/result and the agent final answer.
- 2 Check: A is78, B is95; A fits and B exceeds budget by5.
- 3 Hint: Verify the trace shows actual tool output, then check the arithmetic independently.

Inspect the trace

1

Record request, tool name, input, result, number of iterations and final answer. Mark which claims came from a tool and which from supplied data.

2

Check: A partner can replay why the recommendation was made.

3

Hint: Do not treat hidden reasoning text as a reproducible tool log.

Test unsupported requests

- 1 Ask for current weather, an online purchase and a task that exceeds the configured limit.
- 2 Check: The agent reports a capability boundary or stops; no imaginary execution appears.
- 3 Hint: Use your actual workflow error/stop route to demonstrate the limit.

Evaluate transfer

- 1 Change quantities and budget, rerun, and compare results against a manually computed answer.
- 2 Check: The agent adapts without a new system prompt.
- 3 Hint: Save your own rubric for math correctness, tool use and unsupported actions.

When it goes sideways

01

It answers math without a tool trace

The agent narrated a calculation rather than using the connected tool. Try: Make exact arithmetic tool use part of the rubric and inspect execution data.

02

No tool call occurs

The model or configuration lacks tool-call support. Try: Choose a documented supported model and verify the Calculator connection.

03

It loops or claims unavailable actions

Scope and iteration limits are missing. Try: Set the limit, narrow authority and test an unavailable capability.

Review your evidence

- 1 Supported model and Calculator are actually connected.
- 2 Maximum iterations is configured.
- 3 Tool trace contains input and result.

Review your evidence

- 1 A78 and B95 are independently verified.
- 2 Unavailable weather/purchase requests are handled honestly.
- 3 Changed quantities pass a manual comparison.

What proves calculator use?

- 1 The final answer says I used a calculator.
- 2 The execution trace contains the tool call and result.
- 3 The prompt includes the word tool.

A weather question arrives with only Calculator connected.

- 1 Invent a forecast.
- 2 Explain the unavailable capability.
- 3 Assume a hidden browser exists.

Build with a clear contract

You may use only a connected Calculator tool. Compare $\text{optionA} = 3 \times 18 + 2 \times 12$ and $\text{optionB} = 4 \times 20 + 1 \times 15$ against a fictional budget 90. Return each total, budget difference, tool evidence and a draft recommendation. Do not purchase, browse or send messages. If a tool is unavailable or the task requires another capability, say so. Stop within 4 iterations. A human reviews the recommendation.

[Open copyable prompt](#)

Add an approved-note lookup

- 1 Only approved notes are reachable.
- 2 Unknown facts remain unknown.
- 3 Tool budget and error paths are demonstrated.

Build independently, then teach it back

Your next challenge

Add a read-only notes tool before the calculator. Keep its corpus small and require source IDs. Compare lookup failures with arithmetic failures.

Explain your decisions

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