

1 / 8 A research report

Deep research included

- The job** Does it answer the question you asked, for the reader you named, or an easier one next to it?
- The facts** Open three of its sources. Does each page say what the report says, in the same words and numbers?
- The steps** Which sources did it use, and which kinds did it skip? Is anything surprising backed by two that don't copy each other?
- The changes** What did the summary drop on the way: an "about", "in one study", a year, a sample size?
- The sign-off** Would you quote it to a client with your name under it?

2 / 8 Numbers and spreadsheets

Tables, models, forecasts

- The job** Is it the right data, for the right period, in the right unit: month, currency, thousands?
- The facts** Trace two figures back to the raw data. Do they come out the same?
- The steps** Where did each number come from: a formula, a pasted value, or the AI's own estimate?
- The changes** What did it change in the file? Compare with the last version in the version history.
- The sign-off** Could you explain the biggest number to your finance lead in one sentence?

3 / 8 A customer email

Or any message that leaves the team

- The job** Does it do the one thing this email is for, in the tone this customer expects?
- The facts** Is every name, date, order detail and policy line right, against the record and not the chat?
- The steps** What did it draw on: the customer's own email, your notes, or a guess?
- The changes** What does it promise that you never approved: a refund, a discount, a date, a next step?
- The sign-off** Would you be happy to see it forwarded to their boss? Will a person press send?

4 / 8 A deck

Slides for a meeting or a client

- The job** Does the story lead to the decision this audience has to make?
- The facts** Does every chart match its data, axis and units included? Is every name and quote real?
- The steps** Where did each slide's facts come from? Is the source on the slide or in the notes?
- The changes** If it rebuilt an old deck, what did it update, drop or reword without telling you?
- The sign-off** Could you present every slide without saying "the AI put that in"?

5 / 8 Meeting notes or a transcript

Notes, summaries, action lists

- The job** Does it record what was decided, not just what was said?
- The facts** Are names, numbers and dates right? Does every quote match the recording or transcript?
- The steps** Was it made from the recording, the transcript, or someone's notes? Who is missing from it?
- The changes** Did it smooth over a disagreement, turn a "maybe" into a decision, or drop an objection?
- The sign-off** Would the people there agree with it? Did everyone know it was recorded, as your rules ask?

6 / 8 An agent's run

Tasks it did, on a schedule or once

- The job** Did it finish the task in the brief, and stop where the brief said stop and ask?
- The facts** Is every name, number and date it used real, and read from the place it was meant to read?
- The steps** Is it accessing files or websites you didn't mention? Is the task creeping beyond what you asked for?
- The changes** What did it touch, send and change? Look in sent items, drafts, file history, the calendar, the records.
- The sign-off** Would you let it run the same way next time? If not, move it down a level.

7 / 8 An AI-built app or Site

Small apps, pages, forms, dashboards

- The job** Does it do the one job it was made for, for the people who will use it?
- The facts** Is every number, date, claim and contact detail on it true and current?
- The steps** Where does it read data from, and where does it send what people type in?
- The changes** Who can see it, edit it or share it? What happens when someone changes it later?
- The sign-off** Would you put your organisation's name on it? Who owns it from today?

8 / 8 "It passed a check"

"Verified", "tested", "no errors"

- The job** What exactly was checked: the whole thing, or one part of it?
- The facts** Checked against what: a source, a test, the data, or its own earlier answer?
- The steps** Who or what ran the check? Was it the same AI that made the work?
- The changes** Was the work changed after the check ran?
- The sign-off** Would this check catch the mistake that would hurt you most?

2 / 8 Numbers and spreadsheets**Look first at:** The totals and formulas.**Quick test:** Re-add one total by hand. Trace two figures to the raw data.**Red flags:** Round numbers. Totals that don't match the source. Formulas changed, or replaced by values. A percentage of what?

SORT LIST TRACE HEDGE LOG

A spreadsheet that looks tidy can still add up the wrong column.**1 / 8 A research report****Look first at:** Its sources.**Quick test:** Pick three claims and find each one on its page.**Red flags:** Links that don't open. Quotes you can't find. Sources that agree too neatly. A figure with no year or sample.

SORT LIST TRACE HEDGE LOG

A citation is a place to check, not a check.**4 / 8 A deck****Look first at:** The charts and the numbers on them.**Quick test:** Trace the chart that matters most to its data.**Red flags:** Charts with no source. A neat number with no date. Last year's figures on a "current" slide. Notes that say more than the slide.

SORT LIST TRACE HEDGE LOG

A slide is read in seconds and repeated for months.**3 / 8 A customer email****Look first at:** Who it goes to, and what it promises.**Quick test:** Read it as the person receiving it.**Red flags:** Dates, prices or refunds you never approved. A name spelt two ways. Warm agreement with a complaint nobody has checked.

SORT LIST TRACE HEDGE LOG

Once it's sent, it's the customer's copy, not yours.**6 / 8 An agent's run****Look first at:** Its record and its connectors.**Quick test:** Read the last run from end to end, then look where it acted.**Red flags:** New sites, apps or connectors. A run marked complete that nobody read. Messages sent that nobody approved. A scheduled run nobody owns.

Anthropic's safety page asks (as of Oct 2026): "Is Claude accessing files or websites you didn't mention?"

SORT LIST TRACE HEDGE LOG

A run marked complete is not a run checked.**5 / 8 Meeting notes or a transcript****Look first at:** The decisions and the actions.**Quick test:** Check each action's owner and date with someone who was there.**Red flags:** Actions with no owner. Decisions nobody remembers making. A quote nobody said. "Everyone agreed".

SORT LIST TRACE HEDGE LOG

The notes become the record, so a wrong one becomes the truth.**8 / 8 "It passed a check"****Look first at:** What the check compared.**Quick test:** Ask what it checked against, and look.**Red flags:** "Verified" with no detail. A check of something other than the thing you asked for. The same AI marking its own work.

SORT LIST TRACE HEDGE LOG

When an AI tells you its work passed a check, ask what was checked.**7 / 8 An AI-built app or Site****Look first at:** What it does with data.**Quick test:** Use it as a stranger would: odd inputs, empty fields, a phone.**Red flags:** Real customer data in a test version. A form that sends to an inbox nobody reads. Open sharing. No owner.

SORT LIST TRACE HEDGE LOG

If it works on the first try, try it the way a stranger would.