

The three checks

Reading it over to see if it looks right is the check most people do. It's tuned for the wrong failure — it catches clumsiness, and AI is never clumsy.

These three take about five minutes together. Keep this by your desk for the first month; after that you'll do it without the card.

Check 1 — Find the load-bearing claims

Ask: *If this sentence were wrong, would anyone do anything differently?*

Most sentences fail that test. They're connective tissue, and a wobble there costs nothing. Usually two or three sentences in a document actually matter: the number in the recommendation, the deadline, the reason.

Mark those, and check **only** those properly — open the source, recompute the number, confirm the date.

About two minutes. Catches most factual errors.

Check 2 — Ask what it couldn't have known

Ask: *What do I know about this situation that wasn't in what I gave it?*

This is the one almost nobody does, and it's the one that matters most.

The model produced a reasonable answer for a *generic* version of your situation, because a generic version is all it had. Every gap between that and your actual situation is an error waiting to happen — and it is invisible on the page, because the text is internally consistent.

Things it couldn't have known:

- The history — what was tried before, and how it went
- The politics — who has already objected, and to what
- Last Tuesday — whatever changed that isn't written down anywhere
- The reader — who this is going to, and what they already believe
- The reason the current policy exists

If you find a gap, **give the model the missing context and regenerate**. Don't patch the output by hand — you'll fix the sentence and leave the reasoning wrong.

About one minute. Nothing else catches these.

Check 3 — Recompute exactly one thing

Ask: *Does this one number survive contact with its source?*

Pick a single figure or quotation — at random, not the one you're most sure of — and verify it.

This is a spot check and it works the way spot checks work. It doesn't prove the rest is right. It tells you whether this output is trustworthy **in general**:

- If it's wrong → stop trusting the whole thing and check it properly.
- If it's right → you've earned some confidence, for thirty seconds' work.

Scale the checking to what it costs to be wrong

THE WORK	WHAT IT GETS
Internal note nobody will act on	A skim
Leaving the building, or feeding a decision	All three checks
Money, law, or somebody's employment	All three, plus a second pair of eyes

That last row is worth being blunt about. If the cost of being wrong is a legal exposure or someone's job, "I checked it" isn't a control. Someone who didn't produce it needs to look.

Can a second AI do this for you?

Partly — checks 1 and 3, not check 2. A second model doesn't know your context either, so it has exactly the same blind spot as the first.

If you use one, give it a task it can fail rather than a verdict it can agree with. Not *"is this correct?"* but:

List every factual claim in this document. For each, tell me whether you can verify it, and flag any you believe are wrong.

And give it the source, not just the output. A model checking a summary against nothing is guessing.

Related reading: How to Check AI's Work: You're Looking for the Wrong Mistake