The Fact-Check Habit

> **One-liner:** Catch an AI making a confident mistake — and build a simple verification process you'll use every time.

? Jump in (Tinkerers start here)

Pick a topic you know well — your industry, your hobby, your area of expertise. Something where you can spot errors.

Step 1 — Get a confident answer. Send this prompt:

> Give me a detailed overview of **[topic you know well]**. Include specific facts, statistics, and examples. Be thorough and authoritative.

Read the output carefully. **Find at least one claim that feels off.** It might be a statistic that seems too round, a date that feels wrong, a name that's slightly off, or a causal claim that oversimplifies reality.

Step 2 — Make the AI check itself. Send this:

> Look at your previous response. I want you to fact-check yourself. For each specific claim, statistic, or example you cited:
>
> 1. Rate your confidence (high / medium / low)
> 2. Flag anything you might have fabricated or estimated
> 3. Identify which claims are most likely to be wrong and why
>
> Be ruthlessly honest. I'd rather know what you're uncertain about than have you defend everything.

Step 3 — Verify. Pick the 2-3 claims the AI flagged as lowest confidence. Google them. Were they accurate, close but wrong, or completely fabricated?

Step 4 — Build your check. Based on what you just learned, write a 3-line "verification prompt" you can append to any AI output:

> “ Before I use this, tell me:
>
> 1. Which specific claims are you least confident about?
> 2. What did you estimate or approximate vs. know with certainty?
> 3. What should I verify independently before sharing this?

Save this somewhere you'll see it. Use it as a default follow-up to any AI output you plan to rely on.

? Plan first (Planners start here)

Here's what you're about to do:

1. **Choose a familiar topic** — You need to be able to spot errors, so pick something in your area of knowledge. Don't use an unfamiliar topic — you won't know what to verify.
2. **Generate an authoritative-sounding response** — Ask AI for a detailed, factual overview. The more specific and confident the output, the more likely it contains subtle errors.
3. **Ask AI to fact-check itself** — Use the self-audit prompt to force the AI to rate its own confidence and flag potential fabrications.
4. **Independently verify** — Pick the lowest-confidence claims and check them against reliable sources. Track what was right, close, and wrong.
5. **Create your verification template** — Build a reusable 3-question follow-up that you'll use after any AI output you plan to act on.

"Done" looks like: You've caught at least one AI error, you understand *why* the AI got it wrong, and you have a saved verification prompt you can use going forward.

? Why this matters (Strategists start here)

The community's Ethical Prompting score is 75% — the highest of all five pillars. Most people *know* they should verify AI output, but few have a systematic process for doing so. This exercise closes the gap between awareness and practice by giving you a concrete, reusable tool. The verification prompt you build here becomes a habit — a 30-second step that catches errors before they become problems. At the intermediate level, you'll build a more comprehensive verification

checklist; this exercise establishes the baseline behavior.

Reflection

- What type of error did the AI make — a fabricated fact, a wrong date, or a subtle logical leap? Does the category matter for how you'd catch it?
- Did the AI's self-assessment match what you found when you verified manually? Was it too confident, too cautious, or well-calibrated?
- Will you actually use your verification prompt going forward? What would make it stick as a habit vs. something you forget about?
- □□ *Share a specific AI error you caught with a colleague. Ask them how they currently verify AI output — you may discover they don't.* (Social Learners)

?? Level up

Ready for more? Try EP-Intermediate-01 — where you'll build a comprehensive verification checklist and stress-test it against real AI outputs.

Back to Ethical Prompting & Judgment

Revision #1
Created 2026-03-16 15:48:13 UTC by Admin
Updated 2026-03-16 15:48:13 UTC by Admin