FREE RESOURCE

How to Fact-Check AI Output

The 5-Step Verification Playbook — with ready-to-use prompts

> **"Confidence is not evidence."**

By The Professor · theprofessor.info

BACKGROUND

Why This Matters

AI models predict language, not truth. They generate what a plausible answer should sound like — which means a hallucinated source, a wrong date, or a non-existent study can appear with exactly the same confident tone as a verified fact.

Lawyers have been fined for submitting fake AI-generated case citations. Researchers have chased studies that do not exist. When the mistake is in your work, it belongs to you.

The rule: The higher the consequence, the lower the trust you should place in fluent output.

OVERVIEW

The Verification Framework

Five steps in sequence. Each catches a different type of error — don't skip any.

01	02	03	04	05
EXTRACT	SOURCE CHECK	TRIANGULATE	CHALLENGE	JUDGE
Separate facts from opinion	Verify every citation	Cross-reference 3 sources	Make AI defend its answer	Apply your own reasoning

Extract → Source Check → Triangulate → Challenge → Judge

THE DETAIL

The 5 Steps in Detail

Each step includes a ready-to-use prompt. Copy it, paste it into any AI tool, replace the placeholder, and go.

01 Extract the Claims

Don't trust the paragraph. Break the AI's answer into its individual parts. You're looking for numbers, names, dates, organisations, studies, laws, cases — anything phrased like "research shows". These are the parts that can hurt you.

STEP 1 PROMPT — COPY AND PASTE INTO ANY AI TOOL:

```
Break this output into:
  1. Factual claims
  2. Opinions
  3. Assumptions

List each factual claim as a separate bullet I can verify.
Then rank the factual claims from highest verification
priority to lowest, and briefly explain why.

[PASTE OUTPUT HERE]
```

PRO TIP

Rank before you verify. If the answer has ten claims, don't check all ten with the same intensity. Specific claims (named studies, precise figures) need more checking than general ones.

02 Source Check

Test the easiest failure point first: does the source even exist? Copy any citation the AI has given you, paste it into Google Scholar, and search. If nothing appears, search Google. If it still fails, put the title in quotation marks. A result that is similar but not exact is also a red flag.

STEP 2 PROMPT — COPY AND PASTE INTO ANY AI TOOL:

```
Does this source exist? Provide:
 - Correct title
 - Correct authors
 - Publication year
 - Journal or publisher
 - DOI
 - Working link

If the citation is wrong or cannot be verified, say so
clearly and suggest the closest real source.

[PASTE CITATION HERE]
```

PRO TIP

If you cannot confirm a source quickly, remove it from your draft until you can. Never leave a "maybe" in final work.

03 Triangulate

One source is not evidence. Two is better. Three independent sources is where trust begins. The critical word is independent — agreement across three AI tools is not the same as agreement across three separate institutions or original studies.

STEP 3 PROMPT — COPY AND PASTE INTO ANY AI TOOL:

```
Verify this claim using independent sources.
Return:
 1. Original / most direct source
 2. Best secondary source
 3. Any disagreement or uncertainty between sources
 4. A confidence score out of 10

Do not invent sources. If you cannot verify, say unknown.

[PASTE CLAIM HERE]
```

PRO TIP

Triangulation is non-negotiable for anything high-stakes. If money, reputation, legal exposure, health, or academic integrity is involved, one source is never enough.

04 Challenge the AI

Now make the model defend itself. Paste the original answer back in and ask it to critique its own output. This is where weak answers start to wobble. Truth becomes clearer under pressure. Hallucinations become vaguer under pressure.

STEP 4 PROMPTS — USE ANY OR ALL OF THESE:

```
How confident are you in this answer? Score each claim
out of 10 and explain why.

[PASTE OUTPUT HERE]

---

Assume this answer is wrong. Explain how it could be
wrong, incomplete, or misleading.

[PASTE OUTPUT HERE]

---

List the exact points in this answer that still require
manual verification.

[PASTE OUTPUT HERE]
```

> **PRO TIP**
> Don't just ask for the answer. Ask where the answer could fail. This single habit catches more errors than any other step.

05 Judge

This is the step people skip because they assume the model has already done the thinking. It has not. You are still the final filter. Ask yourself these questions before you use or publish the output:

STEP 5 — QUESTIONS TO ASK YOURSELF:

```
Does this make sense given what I already know?
Would I put my name to this without hesitation?
Does the number look suspiciously neat or round?
Is the quote too perfect, too clean, too convenient?
Is the source oddly difficult to find?
Does this claim serve a narrative a little too well?

Rule: If a claim feels surprising — check harder.
      If a claim feels useful — check harder.
      If a claim feels perfect — definitely check harder.
```

STRATEGY

Using AI to Check AI

One of the best approaches is to generate in one AI tool and verify in another. Competing models expose each other's weaknesses. But remember: the original source always beats the AI summary.

STEP 1
Generate
ChatGPT, Claude, or similar

→

STEP 2
Verify
Perplexity or search-focused AI

→

STEP 3
Confirm
Google Scholar or original source

CROSS-CHECK PROMPT:

```
Compare this answer against independent sources.
Highlight unsupported claims, missing nuance, and
anything that looks hallucinated or stitched together.

[PASTE OUTPUT HERE]
```

QUICK REFERENCE

Complete Prompt Bank

All seven prompts in one place. Replace [PASTE OUTPUT HERE] or [PASTE CLAIM HERE] with your own content.

1 — EXTRACT AND RANK

```
Break this output into:
  1. Factual claims  2. Opinions  3. Assumptions

List each factual claim as a separate bullet I can verify.
Then rank from highest verification priority to lowest.

[PASTE OUTPUT HERE]
```

2 — SOURCE AUDIT

```
For each claim below, identify:
  - The exact source I would need to verify it
  - The best place to search first
  - Any warning signs the source may be fake or out of date

If you are unsure whether a source exists, say so.

[PASTE CLAIMS HERE]
```

3 — CITATION CORRECTION

Does this citation exist? Provide: correct title, authors,
publication year, journal/publisher, DOI, working link.

If wrong or unverifiable, say so and suggest closest real source.

[PASTE CITATION HERE]

4 — TRIANGULATION

Verify this claim using independent sources. Return:
1. Original/most direct source 2. Best secondary source
3. Any disagreement or uncertainty 4. Confidence score /10

Do not invent sources. If you cannot verify, say unknown.

[PASTE CLAIM HERE]

5 — CHALLENGE THE ANSWER

Assume this answer is wrong. Explain the three most likely
failure points: hallucinated source, distorted fact, missing nuance.

Tell me exactly what a human should verify before using this.

[PASTE OUTPUT HERE]

6 — EXPERT DISAGREEMENT

What would a well-informed expert strongly disagree with?
Point to the exact sentence, explain the problem,
and tell me what evidence would settle it.

[PASTE OUTPUT HERE]

7 — SAFE REWRITE

Rewrite this answer so that: unsupported claims are removed,
uncertain points are labelled, every remaining factual claim
is phrased cautiously, no citation included unless certain.

[PASTE OUTPUT HERE]

HOW MUCH CHECKING?

Match Your Effort to the Stakes

Not every task requires the full five-step process. Match your checking level to the consequences.

Task Type	Risk Level	Apply These Steps
Draft email / internal note	Low	Quick scan — read for obvious errors
Social media post	Low–Med	Step 1 + verify any names, dates, numbers
Blog post / article	Medium	Steps 1, 2, 3 — every factual claim
Client deliverable / report	High	Full 5-step process
Legal / academic / medical	Critical	Full 5-step + independent expert review

GOING FURTHER

Tips & Tricks

Watch the confidence of the language
Phrases like “studies show”, “research confirms”, and “experts agree” carry no evidential weight. They are linguistic patterns, not citations. Treat them as red flags, not reassurance.

Numbers need more checking than words
A specific figure (87%, £4.2 billion, 2019) is far more likely to be fabricated or misattributed than a general claim. Always ask: where did this number come from?

Partial matches are still failures
If a source exists but the title, authors, year, or journal are wrong, the citation cannot be used. A distorted citation is as unusable as a fictional one.

Use the model’s uncertainty as a signal
When you challenge an AI and it starts hedging (“it’s possible that…”, “I’m not entirely certain…”), that is useful information. Increase your verification intensity on those points.

Save your prompt chains
Build a personal library of verification prompts you trust. Pasting a single well-crafted prompt saves more time than it costs.

Check the date, not just the source
Even a real study from a real journal can be outdated. In fast-moving fields — technology, medicine, law, economics — ask the AI to flag any claim that may have changed since its training.

AI can help you think faster.
It cannot take responsibility for being right.
That part is still yours.

“Confidence is not evidence.” · The Professor-AI · theprofessor.info