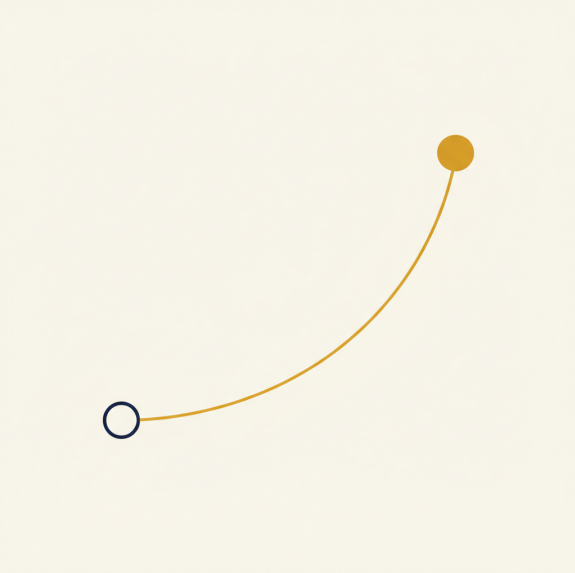
TIER 1 - FOUNDATIONS * V1.0 -- MAY 2026

RECOVERING WHEN AI IS WRONG

The hands-on diagnostic procedure. Spot the wrongness -> diagnose the kind -> re-prompt with the diagnosis -> verify the fix. The recovery skill that turns a dead-end answer into a 5-minute correction.



BY Alex Jahn / Agent Logic

v1.0 -- May 2026 **EDITION**

Anyone who's been using AI for a while and hits the moment when it gives them a confident, fluent, completely wrong answer -- and doesn't know what to do next **AUDIENCE**

15-20 minutes **READ TIME**

Free. Forever. **COST**

Prepared by Agent Logic / alexanderjahn79@icloud.com / theaiguywi.com

CONTENTS

What's in here

You finished Tier 1 -- module 18 of 18. You now have the complete personal-AI toolkit:

SECTION 1

The dead-end moment

"OK, that's clearly wrong -- now what?"

You ask AI a question. The answer comes back fluent, organized, confident. You start using it -- and somewhere in the middle, you realize it's wrong. Maybe you remembered something that contradicts it. Maybe you Googled to verify and the source said the opposite. Maybe a coworker pointed out the gap. Either way: AI was confidently wrong, and you have to figure out what to do next.

Most people, in this moment, do one of three things:

1. **Throw the whole answer out and start over from scratch.** Wastes the work the model already did right.
2. **Ignore the wrongness and use the answer anyway.** This is the dangerous default -- the path that turns AI mistakes into your mistakes.
3. **Argue with the model.** "No, you're wrong, the answer is X." Sometimes works, often produces a fluent apology followed by a different wrong answer.

There's a fourth option, and it's what this module teaches: **diagnose with the diagnosis, verify the fix.** Five minutes of work that turns a bad answer into a recovered correct one.

This is the meta-skill that ties together *Reading AI* and *Use AI* (module 4) -- but as a recovery procedure, not a critical-reading framework. Module 3 told you how to spot wrongness. This one tells you what to do once you've spotted it.

What you'll have by page 13

By the end of this primer:

- The **four kinds** one needs a different fix.

- The **diagnosti**
- The **re-promp**
- The **verificatio**
- **Three worked recoveries** -- a hallucinated reference, a stale fact, a misread question.
- The **honest lim**

> ***Most users treat a wrong AI answer as a dead end. It's not -- most of the time, it's a 5-minute detour. The recovery skill is what separates the people who get value out of AI from the people who keep getting burned.***

SECTION 2

The four kinds of AI wrong

Each kind has a different cause and a different fix

> **The four kinds of AI wrong:**

Kind 1 -- Hallucination. The model generated a confident-sounding fact, citation, quote, or specific that isn't real. Examples: a book that doesn't exist, a "study" with fabricated statistics, a quote from a real person who never said that, a function in a software library that was never written. **Cause:** th
model is predicting the next plausible token; sometimes the most plausible-sounding next sentence isn't tied to reality.

Kind 2 -- Out-of-date. The fact was true at some point during the model's training cutoff, but the world has moved on. Examples: a CEO who left the company last year, a price that's changed, a regulation that was updated, a tool that was rebranded. **Cause:** tra
doesn't.

Kind 3 -- Misread the question. The model gave a competent answer to a SLIGHTLY different question than the one you asked. Examples: you asked about a specific local rule; the model answered about the general national rule. You asked for an option, the model gave the most common option without acknowledging others. **Cause:** an
preference for the most common interpretation.

Kind 4 -- Model limit. The model genuinely doesn't know -- or knows but the question requires reasoning beyond its capability. Examples: hyper-local information AI was never trained on,

freshly-published research, your specific situation that nobody has written about. **Cause:** th
question is outside the model's reliable domain.

The fix is different for each. Misdiagnose, and the fix doesn't help. The whole game is figuring out which kind of wrong you hit.

SECTION 3

The diagnostic procedure

A 30-second checklist to figure out what kind of wrong

Once you've spotted that AI was wrong about something, run through these questions in order. Stop at the first "yes."

> **Diagnostic checklist:**

1. **Is the wrong thing a specific name, citation, quote, statistic, or "specific reference"?** -> If yes, this is a **Kind 1 ha**
2. **Is the wrong thing a fact that was probably true a year or two ago but isn't now?** -> If yes, this is **Kind 2 ou**
3. **Is the wrong thing a competent-but-off answer that addresses a slightly different question than what you actually asked?** -> If yes, this is **Kind 3 mi**
   Section 6.
4. **Is the wrong thing the type of question you'd expect AI to handle, but it just couldn't?** -> If yes, this is **Kind 4 m**

The diagnostic itself is fast. The fix that follows depends on which kind you hit.

The over-arching principle

Across all four kinds, one rule applies: **don't argu**
Re-prompt with the diagnosis instead. Arguing produces apologies and then more confident wrong answers. Re-prompting with the diagnosis gives the model a different cognitive task, which often produces a different (better) result.

SECTION 4

Kind 1 -- Hallucination, and the fix

The signs

You'll spot a hallucination when:

- You search for the cited source/book/quote and can't find it
- The "specific" detail (a name, a function, a fact) returns nothing in a real search
- A coworker / expert tells you the cited reference doesn't exist
- The exact wording feels too clean to be a real quote

The fix

> **Hallucination re-prompt:**

"You said [specific claim]. I tried to verify this and can't find any source. I'm worried that's a hallucination. Can you (1) tell me the exact source you used for that claim -- title, author, URL if possible -- and (2) if you can't produce a verifiable source, retract the claim and tell me what you actually know in this area without inventing specifics?"

That prompt does two things at once: it forces the model to either produce the source (rare, but possible) or retract. Most of the time, the model will retract and give you a more honest answer.

The follow-up move: **never trus**
source. A model that just hallucinated once will probably hallucinate again in the same conversation.

When this is most dangerous

Hallucination is most dangerous when the topic is one where the reader can't independently verify -- research summaries, legal precedent, historical anecdotes, technical specifications. If you're operating in a domain like that, treat AI's specifics as untrustworthy until proven otherwise.

SECTION 5

Kind 2 -- Out-of-date, and the fix

The signs

You'll spot an out-of-date answer when:

- You know from another source that the fact was different recently
- The model references a CEO, product, regulation, or status that you know has changed
- The model uses a price or rate that doesn't match anything current
- The model references "the latest" something that's actually months or years old

The fix

> **Out-of-date re-prompt:**

"You said [specific claim]. I think this might be out-of-date -- it sounds like the situation as of [year you suspect]. Can you tell me what your training cutoff is for this topic, and either (1) flag the parts of your previous answer that are time-sensitive, or (2) tell me explicitly that you don't have current information and that I should verify with a live source?"

The model will usually acknowledge the date issue -- and crucially, will often help you identify WHICH parts of the previous answer are time-sensitive vs which parts are still likely true. The framework parts of an answer (definitions, structures, tradeoffs) tend to age slowly. The specific parts (prices, rates, names, current statuses) age fast.

Then you take the time-sensitive items to a live source. Done.

When this is most dangerous

Out-of-date answers are most dangerous in fast-moving domains: news, market conditions, regulations, technology releases, prices. Build the muscle memory: any answer about a fast-moving topic gets a "what's the cutoff?" check before you act on it.

SECTION 6

Kind 3 -- Misread question, and the fix

The signs

You'll spot a misread question when:

- The answer is competent and confident, but you finish reading it and feel "that's not exactly what I meant"
- The model addressed the general version of your question instead of the specific version
- The model picked one interpretation when you wanted the option list
- You think back and realize your prompt was ambiguous

The fix

> **Misread-question re-prompt:**

"I think I asked the question wrong, or you interpreted it differently than I meant. What I actually need to know is [more specific version of the question]. Specifically: [list the parts of the original question you were ambiguous about]. Try again with that clarified."

Notice the phrasing: "I think I asked the question wrong." This is psychologically easier than "you got it wrong" -- and the model responds better to it. The reframe also forces YOU to be more specific, which is the actual root of the issue.

The misread-question failure is the most common kind of AI wrong, and it's almost entirely the human's fault. The fix is to write better follow-up prompts. The skill compounds: as you do this more, your initial prompts get sharper too.

When this is most dangerous

Misread questions are most dangerous when:

- The topic has subtle variants you didn't think to specify

- Your situation is unusual (the model defaults to common, you needed uncommon)
- The stakes are high enough that "close enough" isn't close enough

SECTION 7

Kind 4 -- Model limit, and the fix

The signs

You'll spot a model-limit failure when:

- You re-prompt with diagnosis and the model just produces the same wrong answer in slightly different words
- The model explicitly hedges ("I'm not sure," "I don't have specific information about this")
- The topic is hyper-local, freshly-published, or extremely specialized
- The question requires reasoning that the model isn't reliably good at (math beyond simple arithmetic, complex logic chains, anything involving real-time data)

The fix

> **Model-limit recognition:**

"I'm starting to think this question is outside what you can reliably answer. Can you confirm: is this a topic where you're confident in your answer, or are you generating plausible-sounding text without strong grounding? If the latter, what kind of human source -- a professional, a current document, a domain expert -- should I be using instead?"

The model will often give you a surprisingly honest answer here. "Yes, this is outside my reliable domain -- you should talk to [X kind of expert]." When that happens, leave the chat and go find that expert.

The point: AI is not a universal answer machine. Recognizing when you've hit a model limit is the difference between AI making you smarter and AI making you confidently wrong.

When this is most dangerous

Model limits are most dangerous when the topic feels like AI should know it. Most people will accept "I don't know" from a tool labeled "encyclopedia." They have a harder time accepting it from a tool that's been confidently answering everything else. Calibrate your expectations: AI is excellent at a wide range of tasks and quietly weak at a few specific kinds. Knowing the difference is the skill.

SECTION 8

Three worked recoveries

Worked recovery 1 -- Hallucinated reference

You asked AI for "a study showing that small businesses with documented SOPs are 40% more profitable." It gave you a confident answer with a specific citation: "the 2019 SBA Small Business Profitability Study by Henderson and Reyes."

You search. You can't find the study. You can't find Henderson and Reyes. You can't find any 40% statistic.

Diagnosis: Kind 1 hallucination.

Recovery:

> *"You said there's a 2019 SBA Small Business Profitability Study by Henderson and Reyes showing 40% more profitability for businesses with documented SOPs. I can't find this study. Was that a real reference, or a hallucination? If it's a hallucination, retract it. If you have a different real reference for the same point, give me that one with the URL or full citation."*

The model retracts. It tells you it doesn't have a specific verified study but that the general principle (documented processes -> better outcomes) is supported in business-management literature. You now know not to use the fake stat. You may or may not have a real one -- either way, you're not propagating the fake.

Worked recovery 2 -- Stale fact

You asked AI about Twitter's content-moderation policies for your social-media research. It gave you a detailed answer about Twitter's trust and safety processes, named several executives, described specific policies.

A coworker reads your draft and says: "These names don't match. Some of these people aren't there anymore. Also, it's called X now, isn't it?"

Diagnosis: Kind 2 out-of-date.

Recovery:

> *"You answered my question about Twitter as if the company is still called Twitter and as if the named executives are still there. I think this might be out-of-date. What's your training cutoff for this topic? Flag which parts of your previous answer are time-sensitive vs which parts are still likely true (e.g., the general structure of how T&S typically works at large social platforms)."*

The model acknowledges the cutoff issue. Confirms several specific facts are stale. Distinguishes the time-sensitive specifics from the structural-process generalities. You now have a useful framework + an explicit "verify these specifics with a current source" list.

Worked recovery 3 -- Misread question

You asked: "What's the law about contractor liability when a subcontractor causes damage on a job site?"

The model answered with a thorough overview of general contractor-subcontractor liability principles in U.S. construction law.

The problem: you needed Wisconsin-specific rules, including the specific WI lien laws and the specific Wisconsin worker's comp implications. The general answer is correct but doesn't tell you what you actually need.

Diagnosis: Kind 3 misread question.

Recovery:

> *"I think I asked the question wrong. I'm a Wisconsin contractor, working in Fond du Lac County. I need the WI-specific version of contractor-subcontractor liability -- including how it intersects with WI lien laws and WI worker's comp. The general U.S. answer you gave is helpful background, but I need the state-specific framework. Same question, narrowed to WI. Note: I'll verify with a Wisconsin construction attorney for anything specific to my actual situation."*

The model gives you the WI-specific layer. You now have what you actually needed. You also flagged that you'll verify with an attorney -- which is the right move for a domain where being wrong has real legal consequences.

> ***Most "AI gave me a wrong answer" complaints are actually misread-question failures. The fix isn't a different model -- it's a sharper second prompt.***

SECTION 9

The verification step

Don't trust the recovered answer either, automatically

Once you've re-prompted and gotten a new answer, don't immediately trust it. Hallucinations breed in long conversations. A model that just retracted a fabricated reference might fabricate a new one in the next response.

The verification step:

1. **Find an authoritative source for the new answer.** A real document, a real expert, a real verified URL.
2. **Check the load-bearing parts** -- the parts where being wrong actually costs you.
3. **If the new answer can't be verified either,** you've hit the model-limit failure. Leave the chat. Find a human source.

The two-pass pattern from earlier modules applies here too: AI for structure, real sources for specifics. Recovery doesn't change the rule; it just lets you stay in the AI session longer before you need to verify.

4

Four kinds of wrong, four different fixes.

Hallucination -> demand source or retract. Out-of-date -> flag time-sensitivity, go to live source. Misread question -> re-prompt with specifics. Model limit -> leave chat, find human expert. The whole skill is in the diagnosis.

SECTION 10

When recovery isn't possible

Three situations where no amount of re-prompting helps:

- **The topic is outside the model's reliable domain entirely.** Hyper-local, freshly-published, classified, niche-specialized. You'll keep getting variations of the same wrong answer. **Fix:** stop trying. Find a human expert.
- **You're emotionally invested in a specific answer.** The recovery skill requires honesty about what you don't know. If you keep arguing with the model because you want a specific answer to be right, you're past the point where AI helps. **Fix:** step a eyes.
- **The model is operating from a fundamentally wrong premise.** Sometimes the model has internalized a misconception so deeply that no follow-up unsticks it. Different conversations, sometimes a different tool. **Fix:** start f the premise correction up front.

Within those limits, the recovery skill turns most "AI was wrong" moments from dead ends into 5-minute corrections. The compounding benefit: as you build the skill, your initial prompts get sharper, the wrongness happens less often, and when it does happen you handle it without losing time.

SECTION 11

Where to go from here

You finished Tier 1 -- module 18 of 18. You now have the complete personal-AI toolkit:

1. What an LLM actually is -- the mental model
2. The 3-question prompt framework -- basic prompting
3. Reading AI output critically -- spotting wrongness
4. When NOT to use AI -- the boundaries

5. AI as a tutor -- deep learning sessions
6. AI for life admin -- the daily applications
7. Make AI know you -- memory and context
8. AI on the go -- voice mode and mobile
9. Free tools, real differences -- the comparison framework
10. AI for big decisions -- the two-pass pattern
11. AI alongside your kids -- parenting framework
12. Reading walls of text -- drag-and-drop summarization
13. AI for hard conversations and personal writing
14. AI and your money
15. AI for daily learning -- the habit
16. Privacy hygiene for AI
17. AI for travel and everyday research
18. **Recovering when AI is wrong**

The complete Tier 1 toolkit. 18 modules. A real foundation.

Tier 2 is next: using AI well at your job. Twelve more modules covering team training, role-specific templates, multi-step workflows, model selection at work, privacy at work, and more.

Get the Tier 2 modules as they drop: theaiguywi.com/training

One email per release. No drip. No spam. Opt out anytime.

If you want this entire 18-module Tier 1 toolkit installed across a team or a family -- the whole curriculum trained in once, not eighteen separate articles to remember -- that's the consulting offer. Same way I run it in my own carpentry business.

> ***Reach out: alexanderjahn79@icloud.com***
>
> A short call. Honest scope. We figure out together if it's a fit.

Closing -- the lock-in line

The recovery skill is the meta-skill. Every other module taught you how to use AI well in a specific domain. This one taught you what to do when the model -- inevitably, occasionally -- gets something wrong. With the recovery procedure, the wrongness becomes a 5-minute detour instead of a dead

end.

You finished Tier 1. Eighteen modules. The full personal toolkit. The compounding starts now if it hasn't already, and Tier 2 picks up at your job whenever you're ready.

18

Eighteen modules.

A complete toolkit for using AI well in your own life -- from "what is this thing" to "what do I do when it's confidently wrong." Tier 1 is built. Tier 2 is next.

Onward.

-- Alex

Agent Logic -- *theaiguyw*

Fond du Lac, WI. This is module 18 of 18 in Tier 1 (Personal). Tier 1 is now complete.