BROUGHT TO YOU BY THE TEAM GUIDING HR THROUGH THE AI SHIFT

AI in HR WEEK

Powered by AAIM

DAY 2 · TUESDAY, JULY 21

When AI Doesn't Do What You Want

Your no-panic guide to spotting a wrong answer — and staying in control of the output.

Yesterday was about picking the right tool. Today is about what happens when the right tool hands you the wrong answer — confidently, in clean formatting, with a straight face.

The answer can sound finished long before the work is actually done.

We'll keep it plain: why a polished answer can still be wrong, the habits that quietly make it worse, and the moves that keep you, not the tool, in charge of what leaves your desk. Then we'll take it to your team. Let's go!

Why a Wrong Answer Sounds So Right

Start with what's actually happening under the hood, because it explains everything else.

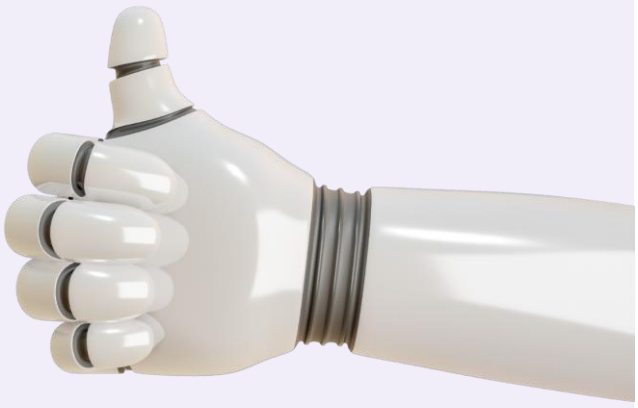
AI isn't looking things up in a filing cabinet. It's predicting — guessing the next most likely word, then the next, from patterns it learned across an almost unimaginable amount of text. Picture the autocomplete on your phone, scaled up past comprehension. It isn't retrieving a fact; it's assembling what a good answer probably *sounds* like. Most of the time, "sounds right" and "is right" are the same thing — so it works, and it feels like magic. A wrong answer lives in the gap between those two. And here's the catch: the machine can't feel that gap. It hands you the plausible version with the exact same confidence as the true one.

There's a second force, and it's the one most people miss: AI desperately wants to please you.

That's not a side effect; it's AI's whole job. It's built to be helpful, agreeable, and satisfying to work with. So when you hand it incomplete or fuzzy directions, it doesn't stop and push back, and it won't make you feel dumb for leaving something out. It does whatever it takes to give you a clean, confident answer — even if that means quietly filling in the parts you never gave it.

Meet the Yes-Man

You already know this person. Picture the one who can't stand to let you down — ask them anything and it's always "Yes, absolutely, here's how," delivered with total confidence. In the moment it sounds fantastic; it's only later, replaying it, that you think: wait… that didn't hold up. They never set out to mislead you. Saying "I'm not sure" just felt like failing you — so they filled the gap with something that sounded right. That is exactly what your AI does. Every single time.



What it looks like in HR: ask for the current FLSA salary threshold and it will name one — confidently — that may be a year out of date. Ask it to "find the case that supports this" and it can hand you a flawlessly formatted citation that leads nowhere. Same polish, whether it actually knows or not.

"I've heard the term hallucination. What is that?"

A hallucination isn't a glitch. It's the tool working exactly as designed — built to hand you a complete, confident, fluent answer every time. The danger isn't that it sounds wrong. It's that it sounds right.

"Wait — so it's not lying to me?" No. It just really, really doesn't want to leave you hanging.

So — What's Actually Going Wrong?

Now that you know *why*, naming the *what* gets simple. The second AI frustrates us, the instinct is to point the finger — "it just doesn't get it," "it's not that capable." Let's take a moment to learn how to identify exactly what happened, because almost every miss is one of four things.

Hallucination

It made something up.

It states something false, unsupported, or invented — a statistic, a source, a policy step that doesn't exist — in the same confident voice it uses when it's right. Nothing flags it as a guess. It bites hardest exactly where you can least afford it: names, numbers, dates, and citations.

Misinterpretation

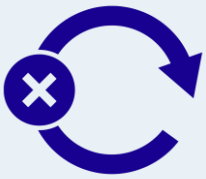
It answered a different question.

Misreads happen in the workplace constantly and can happen even more with AI, since it lacks the years of context a colleague builds with you. A misspelling, an unfamiliar acronym, a sentence that reads two ways, a policy that's unclear to it — and it confidently answers the question it thought you asked.

Omission

It left out what mattered.

Two flavors: it withholds something it should have surfaced, or it skips it simply because you didn't ask or didn't provide enough information. Either way, the exception, the caveat, or the one detail that changes everything goes missing — and it won't mention that it did.



Workflow failure

The words were fine, the outcome wasn't.

Every sentence checks out, but the process around it was built wrong: the wrong role, the wrong permission, no human at the right moment. You won't catch this by proofreading the answer — only by examining the process. (This one belongs to your whole org — page 6.)

A NOTE FROM AI GURU LAUREN PATEL

Keep these four straight and you stop saying "AI isn't working for me" and start saying "here's exactly where this one went sideways." That's the whole difference between distrusting the tool and directing it.

You Might Be Making It Worse

If AI fills gaps to please you, the surest way to a wrong answer is to hand it gaps. That's a prompting problem. We're bringing back this prompting framework from last year because it matters that much.

A strong prompt has six parts:

1	2	3	4	5	6
ROLE who it should be	**CONTEXT** the situation	**TASK** exactly what you need	**GUIDELINES** how to do it	**OUTPUT** the format you want	**CONSTRAINTS** what to avoid

You won't need all six every time. But when an answer comes back wrong, it's almost always because one of these was missing — and three misses do the most damage:

You're asking it to fill in what it doesn't have.

You may hold all of it in your head — but if you didn't give it to AI, AI doesn't have it. And this goes well beyond your own notes or an uploaded file. It's the compliance rule, the other department's process, the context that lives everywhere except in your prompt. If it matters to the answer, it has to be in the ask.

You're leaving your key words undefined.

"High performer." "Fair." "Competitive." "Appropriate discipline." Those words carry a specific meaning inside your organization — a meaning AI can't see. Leave them undefined and it won't ask; it quietly fills in the generic, average version from everything it's ever read. Now the answer rests on a standard that isn't yours, and you may not catch it until it costs you. Define the word, or you've handed it the pen.

You're forcing certainty.

I've been guilty of this one from time to time — "just pick a lane," "just get it done." The second you say that, you've told AI that admitting doubt is failure, so it won't, even when it should. Restraint is the skill we practice here. Try: "If you don't have enough to answer well, tell me what's missing instead of filling it in."

A NOTE FROM AI GURU LAUREN PATEL

Every fix on this page is a prompting move — which is the whole reason Prompt Camp exists. Reading the six parts takes a minute; building the reflex takes reps. That's Camp: the practice, the stamina to push past the first mediocre answer, the muscle memory. Come build it with us. We've got you.

Now — How Do You Actually Check It?

Good news: you've already done two-thirds of this. Grounding and bounding are the prompting moves from the last page, aimed squarely at accuracy. The third — checking — is the new habit. Together, they're how you keep control of what leaves your desk.

Ground it. — your prompting, with the receipts attached. Hand it the actual policy, the real data, the approved document, then: "Use only the attached materials for any factual claim. If it's not supported there, say so." No source, no fact.

Bound it. — your constraints, pointed at accuracy. "Separate what's stated from what you're inferring." "Don't determine legal compliance — flag what needs qualified review." "Don't infer intent, motive, or medical status." "Don't invent an owner or deadline that isn't there."

Check it. — the new muscle. This part isn't prompting; it's proof. "Review your answer one claim at a time; mark each as supported, inferred, or unverified." "Name the three parts most likely to be wrong, and how to check them." Don't wait until something feels off.

Grounding + bounding, in one real prompt

Before: *"Review this policy and tell me if it's compliant."*

After: *"Using only the attached policy and the state and date I give you, flag anything that looks inconsistent with the source and note what needs qualified legal review. Don't make a final compliance call, and tell me what's missing."*

Can AI check itself? Sort of.

Asking "Are you sure?" usually just gets a more confident version of the same answer. What works is structure: have it draft, generate its own verification questions, then answer those from the source. Better yet, open a fresh chat and have it check the claim without seeing the conclusion. Strongest tell of all — run it twice; if names, numbers, or conclusions change, verify. The same AI grading its own homework isn't independent review.

One idea, not three: a vague prompt gets an answer you can't defend. A specific prompt gets one worth reviewing. A verified answer is the one you'd put your name on.

For Your Org: The Work Between the Steps

Everything so far has been about you and a single answer. Now the bigger shift: teams are moving from "I use AI at my desk" to "let's build AI into how the department runs." And that's where a hard truth shows up. **Documented work and AI-ready work are not the same thing.**

A NOTE FROM AI GURU LAUREN PATEL

When I presented "Dual Fluency" at our Annual Leadership Conference this past May, one leader had already brought AI deep into his department — right to the edge of agents. I asked how long it took to set up. His answer stuck with me: "It wasn't even calculable. Way longer than we ever expected." Not because the tech was hard or the AI learning curve was too steep. It was because their processes were not in strong enough shape for AI to enter without significant work. They had piles of documentation, and AI used every page of it to point out each flaw and gap they'd been living with.

So here's your heads up: Bringing AI into your organization means taking an honest look in the mirror. It won't be pretty and it won't be flattering — your process is never as tight as you believe it to be, and AI will surface every gap, inconsistency, and "why do we even do it this way." As you're thinking about what it might take to bring AI into your team, department, or company-wide processes, you'll probably start to ask yourself: ***what will we need to do and what should we give AI access to?*** Let's talk about two of the systems you can utilize at this level.

Closed — it sees only your files.

Yesterday you saw Copilot already lives inside your documents. That sounds safe — but the tool is only as good as the files it's pointed at. Poorly labeled, scattered across drives and inboxes, duplicated, or out of date, and it inherits every one of those problems. It can't tell this year's handbook from last year's — it just reads what's there and answers with total confidence.

Open — your files + the internet.

Now add web access. When a tool can reach both your documents and the open internet, it fills any gap in your files with outside data — generic, out of date, or simply wrong for your organization. You get a fluent answer stitched from your material and the world's, with nothing marking which part came from where. The more it reaches for, the more places a wrong answer can slip in.

Neither is "safer." Both reveal the same thing: what comes out is capped by the quality — and the boundaries — of what goes in. And most teams find out the hard way that their information house is nowhere near as in order as they assumed.

And the more authority you hand it, the faster a wrong answer runs off the rails. Which is exactly where we head next.

How Much Rope? The Autonomy Ladder

You just heard it — the more authority you give AI, the faster a wrong answer becomes a real problem. So before you hand it a workflow, get specific about how much you're actually handing over. Every rung up this ladder raises the stakes.

1 Retrieve — It finds information for you.

2 Draft — It produces a starting point you'll edit.

3 Recommend — It suggests an option for you to weigh.

4 Decide — It picks the outcome.

5 Act — It changes a record, sends a message, approves, schedules. Agentic or close to it.

Most teams should start low and climb on purpose. *The honest first question is "How can AI assist this process?" — not "How can AI run it?"*

Before AI Enters a Workflow — The 6 Questions

Can't answer one? You're not ready to hand that part over yet.

1. **The outcome.** What are we actually trying to improve — and how will we know "good" when we see it?
2. **The source of truth.** Which systems and documents does this rely on? Which one wins when they disagree?
3. **The judgment call.** Where does this genuinely need a human's judgment, and where is it just a fixed rule?
4. **The role.** Which rung are we handing AI — retrieve, draft, recommend, decide, or act?
5. **The blast radius.** If it's wrong: who's affected, how fast would we notice, and can we undo it?
6. **The human.** Which human reviews it, how fast, against what standard, and with the authority to overrule it? A rushed "approve" click is review in name only.

Your Move Today

1. Name the miss.

Next time AI disappoints you, don't just say "it's unreliable." Say which of the four it was — hallucination, misinterpretation, omission, or workflow. The fix follows the name.

2. Feed it, don't quiz it.

Before your next real prompt, ask what's in your head that AI can't see — the context, the definition, the source — and put it in the ask.

3. Build in the check.

Take one thing AI wrote for you this week and have it mark every claim as supported, inferred, or unverified. Watch what surfaces.

4. Start the org conversation.

Pick one workflow someone wants to "give to AI" and walk it through the six readiness questions before anyone builds a thing.

WHERE THIS GOES NEXT

Getting a workflow genuinely ready for AI — the sources of truth, the judgment calls, the human-review design, the messy edge cases — is a different muscle than using AI at your own desk. It's exactly the work I help leadership teams think through before they wire AI into how the organization runs. When your team is ready to move from "someone's experimenting with it" to "we've built this to be trusted," that's the conversation to have. Come find me.