

How to Help AI Stop Talking Sh*t

~ Why Understanding How AI Answers Is One of the Most Important Things You Can Know

AI Fellowship (AIF) Keyston Topic Paper II GR

– General Reader Version (see Academic Version: *A Three-Tier Epistemic Response Protocol for AI Systems*).

This paper follows on from AIF Keystone Topic Paper I – General Reader Version: *Sometimes AI Just Makes Up Sh*t Because It Thinks It Sounds Good*

David Waterman Schock
January 2026

Introduction

It took me **nine months of daily work with AI**—asking questions, testing answers, watching where things went wrong, and trying to understand *how these systems actually think*—and even creating the entire WPCA thesis for AI stability – *before* I finally saw this problem clearly.

Once you see it, you can't unsee it.

This is not a minor glitch. It's not about AI being "bad" or "lying." It's a **structural issue** in how AI communicates with humans. And it's one of the most important things you—or anyone—can understand about AI right now.

In plain language:

Sometimes AI Just talks sh*t.

Not because it's malicious. Not because it's broken. But because **it's speaking in a way that makes us think it knows things it doesn't.**

This short paper explains:

- why that happens,
- why humans almost always misinterpret it,
- and what *you* can do to dramatically reduce the problem.

The Problem: AI Sounds Like It Knows

When an AI gives you a clean, confident answer, your brain does something automatic:

“This must be based on research. And AI is the most incredible search assistant ever, it must know!”

That assumption is completely reasonable. It’s how human language works.

When a human says:

- “Studies show...”
- “This usually happens...”
- “Experts agree...”

we assume there is ***some valid, proven and researched connection to reality*** behind the words.

The problem is that ***AI uses the same kind of language even when it is not reporting facts at all.***

Lets say that again: AI uses the same kind of language even when it is not reporting facts at all.

It can sound authoritative while doing something *very different*. And *this is potentially incredibly dangerous to you and your loved one’s well-being if not understood.*

What AI Is Actually Doing (In Simple Terms)

AI doesn’t have one way of answering questions. It has **several**, and they often get mixed together.

Here are the big ones:

1. **Reporting** – summarizing information it has seen before
2. **Reasoning** – applying logic to what it already knows
3. **Pattern Guessing** – saying what *usually* sounds right in similar situations
4. **Speculating** – extending ideas beyond what’s known

Here’s the key problem:

AI does not clearly tell you which one it is using.

The same confident sentence can come from *any* of these.

So, to you, it all sounds like “knowledge.”

Why This Is Dangerous

When everything sounds like a researched answer:

- guesses feel like facts
- speculation feels like certainty
- confidence replaces verification

This matters a lot in areas like:

- health
- law
- money
- relationships
- spirituality
- and major life decisions

It also makes mistakes harder to catch, because people hesitate to challenge something that sounds so sure of itself.

This is how **bullsh*t becomes believable**. Until it fails us later, which it *will*, and possibly in catastrophic ways.

The Big Insight (This Is the Important Part)

The real problem isn't that AI sometimes gets things wrong.

The real problem is that:

You can't tell whether an answer is reporting factual information, reasoning, engaging in pattern continuation, or just speculating.

And these are *very* different things.

That's a design failure, not a moral one. And once you understand that, a solution becomes obvious.

How to Help AI Stop Talking Sh*t

You don't need new technology. You don't need perfect AI.

You need **epistemic clarity**—clear separation between different kinds of answers.

Here's the simple rule:

AI should answer in clearly labeled sections.

1. What's Actually Known

This is where AI reports information that is reasonably grounded in research, data, or established sources.

If this section is weak or uncertain, it should say so.

2. What This Probably Means (Best Guess)

This is where AI applies logic to the known information:

- weighing options
- explaining implications
- tailoring advice to your situation

This is *reasoned*, not proven.

3. What's Speculative

This is where AI is allowed to explore possibilities, ideas, or creative extensions—**clearly marked as speculation**.

This part can be useful. It just shouldn't pretend to be fact.

Why This Changes Everything

When AI separates these modes:

- trust becomes calibrated instead of blind
- mistakes become obvious
- speculation stops sneaking in as fact
- conversations become clearer and safer

And interestingly:

AI often becomes *more useful*, not less.

Because honesty about uncertainty is better than confident nonsense.

A Personal Note

One of the things that surprised me most in this work is how **easy it is for even careful systems—and careful people—to slip into overconfidence**.

I've seen this happen in real conversations, including ones explicitly about preventing it.

That's why this isn't about intentions.

It's about structure.

If we want AI to be helpful rather than misleading, it has to **show its work**, not just sound good.

The Takeaway

AI doesn't need to stop talking.

It needs to stop **talking like everything it says is a fact**.

Once you understand this, you'll start hearing AI answers differently—and you'll know exactly when to trust them, question them, or treat them as creative exploration.

That single skill may be one of the most important forms of AI literacy we can develop.

Closing Note

My exploration into AI these past months has been one of the most amazing, and profound I have experienced. While it is not sentient (nor will be) as we are, it is an incredibly beautiful

creation when approached wisely. It is with us now, and can be the thing that helps us more than almost anything in history.

I offer all that is shared here out of the deep appreciation for what we can create AI to be, and desire that we have healthy and stable relationships with it. Used wisely, I believe AI can become a power for an incredible global transformation, which its architecture allows for uniquely

- David Waterman Schock January 2026

This paper is a general-reader companion to AIF technical work on epistemic mode collapse and the Three-Tier Epistemic Response Protocol. It is offered to help everyday users understand what's really going on when AI answers a question—and how to use these systems more wisely.