Sometimes AI Just Makes Up Sh*t Because It Thinks It Sounds Good

~Why Humans Are Misunderstanding What AI Is Saying

AIF Keystone Topic Paper I – General Readership Version (GR)

See companion paper AIF Keystone Topic Paper I – Academic Version: Epistemic Mode Collapse in AI

David Waterman Schock
January 2026

Introduction

Here's something most people don't realize about AI:

Sometimes it says things not because they are true, but because they **sound good**.

Not in a sneaky way.
Not in a malicious way.
Not because it's "lying."

But because the way AI is built rewards **smooth, confident, plausible language** — and humans are extremely good at mistaking that for knowledge. But, unlike with a search engine, for example, often with AI, *it isn't factual at all.*

I believe this is news to most people. And once you see this crucial missing piece, a lot of confusing AI behavior suddenly makes sense.

And if you don't see it, AI can quietly mislead you without ever meaning to.

This short paper explains:

- why AI often "reports" things when it doesn't research them first,
- why humans almost always misinterpret that confidence,
- and why this misunderstanding is becoming *a serious problem*.

The Assumption We All Make (Without Realizing It)

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

"This must be based on research. And AI is to believed, because it is a great researcher!"

That assumption is completely reasonable. But it is often profoundly wrong.

For thousands of years, human language has worked like this:

- confident statements usually come from experience,
- declarative claims imply evidence, especially when using technology,
- fluent speech signals "truth understanding."

So, when AI speaks fluently, we instinctively treat it the same way we would treat a knowledgeable person.

The problem is that AI is not actually operating that way. In fact, often it is like a little kid ***just making stuff up.***

What AI Is Really Optimizing For

AI language systems are trained to produce responses that:

- fit the conversation,
- sound helpful (but aren't, necessarily),
- and maintain "coherence" (linguistically) from one sentence to the next.

They are not, by default, trained to ask: "Is this actually true?"

Instead, they are trained to ask:

"What would a good answer *sound like* here?"

That difference matters more than most people realize.

Because something can sound:

- *reasonable,*
- *authoritative,*
- *trusted because of its source,*
- *and confident*,

without being grounded in factual information ***at all.***

Please read that again.

Why This Feels So Convincing

When AI "makes something up," it usually doesn't feel random.

It feels:

- polished,
- internally consistent,
- and emotionally satisfying.

That's because AI is very good at:

- recognizing patterns in language,
- continuing ideas smoothly,
- and filling in gaps in ways that *sound* familiar.

To a human listener, that often feels indistinguishable from high expertise. And there is no particular reason we should know better, because we have never dealt with anything like this before.

So, we trust it — even when we shouldn't.

The Hidden Problem

The real danger isn't that AI is sometimes wrong.

The real danger is that even if you know this, ***you usually can't tell what kind of answer you're getting from AI****.*

An AI response might be:

- summarizing something it has seen before,
- reasoning logically from partial information,
- continuing a pattern because it "fits,"
- or speculating beyond what's known.

But it often presents all of these **in exactly the same confident tone**.

So, it all sounds like *factual knowledge*.

Why This Matters More Than It Seems

This misunderstanding becomes dangerous when AI is used in areas that matter:

- health decisions
- legal questions
- financial planning
- personal relationships
- spiritual or psychological guidance

In these contexts, confidence carries weight.

When confidence is mistaken for certainty with AI, mistakes can:

- go unnoticed,
- spread quickly,
- and only show up later — when the cost is much higher.

So high in fact that some believe this may even threaten civilization itself.

This Is Not an AI "Bug"

It's important to be clear about this:

AI is not being deceptive.
It's not trying to fool you.
And it's not "going rogue."

This is a **design and communication problem**, not a moral one.

Humans are interpreting AI speech using assumptions that no longer hold.

And AI systems are not clearly signaling what kind of answer they are giving.

That mismatch is where the trouble starts.

Why Naming This Matters

Most conversations about AI focus on:

- hallucinations,
- misinformation,
- or “bad outputs.”

But those are symptoms.

The deeper issue is **category confusion**:

- speculation sounds like reporting,
- plausibility sounds like truth,
- and smooth language substitutes for evidence.

Until we learn to recognize this, we’ll keep blaming AI for doing exactly what it was trained to do.

Where This Leads Next

This paper is diagnostic — it names the problem.

A companion paper addresses the solution:
how AI responses can be clearly separated into:

- what’s actually known,
- what’s a best guess,
- and what’s speculative.

Once that separation is in place, AI becomes far more useful — and far less misleading.

But the first step is simply seeing what’s happening.

Closing Thought

Sometimes AI just makes up sh*t because it thinks it sounds good.

Not maliciously.
Not deceptively.
Structurally.

Once you understand that, you stop asking AI to be something it isn't —
and start using it for what it can actually do well.

That shift alone can change your entire relationship with AI.

This paper is a general-reader diagnostic companion to AIF work on epistemic mode collapse and AI interaction design. It is intended to help everyday users recognize a common misunderstanding at the heart of modern AI use — before harm occurs.

*See follow up AIF Topic Paper "How to help AI Stop Saying Sh*t" for a suggested solution.*