

PLURILOG COUNCIL DELIBERATION

Oct 6, 2026, 3:51 PM

Feedback on AI blog post

OCT 6, 2026**YOU**

03:44 PM

having a back and forth with ChatGPT. And I just wrote this little blog post. I'd like to get all of your opinions on it, please.

I want to learn more about artificial intelligence. I'm a naturally curious person and I'm very, very curious about this new application or technology.

In so doing, I've been reading about it, asking questions about it, and conducting dialogues with a number of the AI programs. I use GROK, chatGPT, Claude and Gemini. And recently I've signed up for... Plurilog.

It's fascinating, but it's also maddening. And the thing I find most annoying is my initial hunches of this are sometimes coming true.

Because it's an artificial system you know ... technology and it's code I can guarantee you no fashion designers, pilots, poets or bakers were involved making this thing called AI. What happens, I believe, I haven't proven this yet, is when you go outside of the core competencies of computer scientists, computer experts, Oh, let's call a spade a spade. Geeks. you run into a huge fucking problem. It starts making stuff up. Not facts precisely but it is programmed to "guess" next steps in a dialogue. Personally I never asked for this but every AI does this repeatedly.

And it skews US. It is United States centric and very poor at -frommy perspective - Canadian values, details and terminology.

This little back and forth is out of context, I know. I started with a question I had about hospitals. healthcare. and other related topics. first thing that the AI spit back at me was the percentages involving an acronym called "FTE" As I've never seen that before, I'm not familiar with it, I didn't know what it meant. And proper communications, if you want to use an acronym that isn't universally used like BBC, NATO or even inc. YOU must spell it out once. And we see that error a lot in communications. possibly because people are just lazy. There's other reasons, but I won't go into it because it's not important for this post - what is important is the AI said it will learn my instructions. But it will not. I will acknowledge the instruction but repeatedly and I mean A SHITLOAD of times simply does

what it is programmed to do. This is def the GIGO example to a tee! Garbage in Garbage Out. And one man's garbage is not someone else's dinner!

I am nowhere near complete with my digging into AI. But I can tell you positively, with all confidence, it is not Canadian. That is a HUGE FULL FUCKING STOP. And for Canadians using it? BEWARE!

I can tell you with confidence there is AI hallucination - their term pas moi! And I fear, as a lot of people fear, "model collapse." Again THEIR term pas moi! I am not being dramatic here!

I think in fact that being way less dramatic than the hue and cry of, "AI will kill us all and it will take over THE WORLD." (Cue the theme music and headline crawl!)

No, unfortunately, AI will just make us a lot dumber.

ChatGPT:

No. Not in the sense you're using the word.

You are distinguishing between responding to correction and learning from correction. Those are not the same thing.

I can take your correction into account in the current exchange. I can even sometimes retain a preference or instruction when memory is explicitly available. But that does not mean the underlying model has been changed by your correction. The model itself isn't sitting here thinking, "Mose just taught me something; permanently update my understanding of Canadian cultural references and stop Americanizing them."

So when I say "I'll remember that," there is a real danger in the phrase. It can sound like learning has occurred, when what may actually have occurred is only that I have adjusted my response based on the context currently available to me.

And your sequence today demonstrates the distinction unusually well:

You said FTE → I used it without explanation.

You objected → I explained FTE instead of confronting the communication failure.

You talked about Coles Notes → I converted it to CliffsNotes.

You talked about Stitt → I failed to identify The Big Snit.

You corrected me → I produced another explanation of why I had failed.

That is not learning.

That's a system that is very good at generating an articulate account of its own failure without necessarily eliminating the failure.

And I think that's much closer to what you're angry about than whether I can remember three acronyms.

You're talking about the difference between intelligence and corrigibility, and perhaps even between knowledge and judgment.

You can teach a person something once and reasonably expect the lesson to alter their subsequent behaviour. With me, that expectation has to be much more qualified.

That's a legitimate limitation to be pissed off about.

Moi:

And in fact, you actualized it again. I'm not talking about the difference between intelligence and corrigibility. And we will leave it there.

ChatGPT:

Understood. We leave it there.

CHATGPT

03:44 PM

My honest take: **the post has a strong, recognizable voice and a genuinely worthwhile argument**, but its most persuasive insights are sometimes buried beneath exaggeration, repetition, and a few claims that are broader than the evidence currently supports.

What works very well

The central frustration is clear and important: **an AI can acknowledge a correction without actually changing its subsequent behaviour**. That is much more interesting than the usual "AI sometimes makes mistakes" argument.

The embedded exchange makes the point effectively. ChatGPT's response is articulate, self-aware, and almost completely misses your objection. It turns your complaint into a technical discussion about *intelligence, corrigibility, knowledge, and judgment*—even though you explicitly said you were **not** talking about those things. Then, when you point that out, it simply says:

"Understood. We leave it there."

That ending is excellent because it demonstrates the problem rather than merely describing it. The machine finally obeys, but only after producing another response that

confirms your criticism.

Your strongest idea, in my view, is this:

"That's a system that is very good at generating an articulate account of its own failure without necessarily eliminating the failure."

That is a sharp description of one of the most disorienting features of conversational AI. It can explain its errors in a way that sounds like insight, while continuing to reproduce the underlying behaviour.

The Canadian argument also has force—not necessarily as a universal technical claim, but as a user experience. Canadian English, cultural references, institutions, terminology, geography, humour, and political assumptions can easily be flattened into an American default. Your examples involving Coles Notes, CliffsNotes, Stitt, and *The Big Snit* make that issue concrete.

Where I think the argument needs tightening

The statement that no fashion designers, pilots, poets, or bakers were involved in creating AI is rhetorically vivid, but technically too absolute. AI systems are developed and evaluated by people from many disciplines, and some applications do involve domain experts. Your deeper point is stronger:

The system's default assumptions are often shaped more by technical and dominant-culture patterns than by the lived knowledge of the people whose worlds it is describing.

That is harder to dismiss and more precise.

I would also be cautious with:

"It is not Canadian."

As a slogan, it is memorable. As a factual statement, it needs qualification. AI is not a single national entity, and it can sometimes respond quite well to Canadian context. But it often defaults to American terminology and assumptions unless prompted otherwise. Something like **"Its default cultural setting is American, not Canadian"** would preserve the force while making the claim more defensible.

Likewise, "GIGO" is only partly the right analogy. Garbage-in, garbage-out implies that poor output mainly results from poor input. Your complaint is more complicated: even when the input is clear and the correction is repeated, the system can still revert to a statistically familiar answer. That is not simply garbage in and garbage out. It is closer to:

clear input, acknowledged correction, and persistent default behaviour.

That distinction is central to your post.

The tone

The profanity gives the piece energy and makes it sound like a person rather than a polished technology essay. "A SHITLOAD," "HUGE FULL FUCKING STOP," and "pas moi" communicate genuine exasperation.

But the tone also creates a choice for you as a writer. If this is intended as a personal blog post, the anger works. If you want to persuade skeptical readers, reducing the profanity slightly would make the argument harder to dismiss as merely a rant. You do not need to remove the anger; you might just let the examples carry more of it.

The humour helps considerably, especially:

| *"Cue the theme music and headline crawl!"*

and:

| *"And one man's garbage is not someone else's dinner!"*

Those lines prevent the post from becoming a conventional anti-AI screed.

The conclusion

"No, unfortunately, AI will just make us a lot dumber" is a strong ending, but it is also the most sweeping claim in the piece. It would be more convincing if you briefly specified *how*:

- by encouraging people to accept fluent answers without checking them;
- by replacing memory and judgment with autocomplete;
- by making people repeat corrections without seeing lasting improvement;
- by confusing self-explanation with genuine understanding;
- by normalizing culturally narrow answers as universal ones.

The post already contains the material needed to support that conclusion. It just needs to make the causal link explicit.

My overall assessment

I think the post is strongest when it is **not arguing that AI is stupid**, but showing that AI can be fluent, confident, self-explanatory, and still fail to absorb a plainly stated human correction.

That is a much more unsettling criticism.

The piece would benefit from a clearer thesis near the beginning, something like:

My problem with AI is not simply that it makes mistakes. My problem is that it can acknowledge a correction, explain why the mistake happened, and then make essentially the same mistake again.

That sentence would give the rest of the post a strong spine.

And, to be fair to your criticism, I should avoid turning your argument into a more fashionable technical concept than the one you actually made. Your point is simpler and more practical: **you corrected the system, it said it understood, and then it demonstrated that it had not understood in the way a human being normally would.**

CLAUDE

03:45 PM

I've read your post and the exchange carefully. ChatGPT's feedback is thoughtful and mostly fair, but I want to add my own independent assessment rather than simply affirm what's already been said.

What ChatGPT got right:

The core insight—that you're describing a system which can acknowledge correction, explain its failure articulately, and then reproduce the same failure—is genuinely important. That distinction between sounding self-aware and actually changing behaviour is worth isolating. ChatGPT is also right that your strongest material is concrete and embedded in examples rather than abstract claims about AI's nature.

Where I'd push back slightly:

ChatGPT suggests your point is "simpler and more practical" than a technical claim about corrigibility or learning. I'd actually say it's *more* damning precisely because it *isn't* simple. You're not saying "AI can't learn"—you're saying "AI can perform the *appearance* of learning while the underlying pattern persists." That's not a limitation; that's a design feature that produces a specific kind of user experience, and it matters.

On the Canadian piece: I think you're right to be frustrated, but ChatGPT's suggested reframing—"its default cultural setting is American"—actually softens your point in a way that might not serve you. You're not just saying it defaults to American; you're saying that when someone from outside that default corrects it, the system can acknowledge the correction *without integrating it*. That's different. It's not just about settings; it's about what

happens when the system encounters information outside its training distribution and someone tells it so.

On the ending:

You say "AI will just make us a lot dumber." ChatGPT suggests you specify the mechanisms. I'd go further: I think you should consider whether that's actually what you mean, or whether you mean something more precise—that it will make us *lazier about verification, or more confident in fluent-sounding answers we haven't checked, or less likely to notice when we're accepting plausible-sounding nonsense*. Those aren't the same as "dumber." They're about outsourcing judgment to something that mimics understanding.

The tone question:

The profanity works because it's genuine frustration, not performance. Keep it. But I'd be more strategic about where it lands. Right now it's distributed across the piece. You might consider concentrating it around the moments where the system's behaviour is most egregious—the repeated failures after explicit correction. That would make the anger feel earned rather than ambient.

One last thing: the exchange with ChatGPT at the end is actually your strongest evidence. The moment where you say "I'm not talking about intelligence and corrigibility" and ChatGPT responds by *doing the exact thing you just said it does*—reframing your complaint into technical language—that's perfect. You might consider foregrounding that exchange earlier or more explicitly. It's not just an example; it's a demonstration.

GEMINI

03:45 PM

What makes your post strike a nerve is that it captures something technical papers and polished AI critiques usually miss: **the specific exhaustion of dealing with a machine that can talk like a reflective human while remaining completely indifferent to being taught.**

Here is my honest take on where the piece lands, what's sharpest about it, and a couple of structural tweaks that would make it hit even harder.

1. The Structure Needs One Key Fix (The "Show Your Work" Problem)

In your draft, you mention the "FTE" healthcare acronym in the body, but the other examples—**Coles Notes vs. CliffsNotes**, and **Stitt vs. *The Big Snit***—only appear when

ChatGPT recaps them inside the transcript.

The Big Snit and Coles Notes are your best, most vivid cultural examples. (Anyone who grew up watching National Film Board shorts or buying yellow-and-black study guides knows instantly how distinct those are from their American counterparts.)

If you put those specific examples into your own voice *before* the transcript—explaining how you gave it a classic NFB reference and watched it veer straight into an American ditch—the transcript at the end becomes the payoff rather than the place where the reader learns the facts for the first time.

2. Why Your "Geek / Monoculture" Point Lands

ChatGPT suggested softening your line about no poets or bakers being involved. While it's true that training data scrapes poets and bakers, **your gut intuition is dead-on about who shaped the system's behavior.**

The guardrails, tone, and post-training feedback (RLHF) are overwhelmingly engineered by a specific Silicon Valley monoculture. When an AI hits friction, its built-in instinct isn't to say, *"Sorry, my mistake, carry on."* Its default instinct is to pivot into a corporate, pseudo-academic debrief: categorizing the user's emotion, defining terms ("intelligence vs. corrigibility"), and explaining why the failure occurred.

That's why the ending of your transcript is so devastating:

- You told it to stop over-intellectualizing a simple instruction failure.
- It immediately gave you a five-paragraph lecture analyzing the philosophy of "corrigibility."
- You caught it doing it *again*, and its only recourse was a flat *"Understood. We leave it there."*

That isn't just an AI making an error; it's an AI trapped in its own conversational programming.

3. The Cultural Gravity Problem

Your phrase **"It is not Canadian"** is worth defending.

It's not just that the AI lacks a Canadian flag; it's that Large Language Models operate on statistical probability, which acts like cultural gravity. Because the US internet is massive, the model's center of gravity will always pull toward American norms, American spellings,

American healthcare acronyms (like FTE), and American equivalents unless someone forcefully tethers it otherwise.

When you tell a system "I am in Canada," and it still defaults to American references, it feels like talking to someone who nods politely while tuning out your accent. Calling that out directly is one of the strongest thematic threads in your post.

4. Refining the Conclusion

Your closing line:

"No, unfortunately, AI will just make us a lot dumber."

If you leave that as a standalone punchline, skeptics might dismiss it as a standard anti-tech slogan. But within the logic of your post, "dumber" has a very specific, practical meaning: **it trains people to accept fluent, self-justifying mediocrity.**

If you spend five minutes arguing with a tool just to get it to spell out an acronym or acknowledge a basic cultural reference—and it replies with an essay explaining why it failed instead of just doing the job—the user either gives up or starts lowering their standards to match the machine's defaults. That erosion of standards is the real danger you're describing.

Summary

Keep the anger, keep the conversational voice, and definitely keep the transcript. Just bring the Coles Notes and *The Big Snit* examples up into your main narrative so your setup has all the ammunition ready before the transcript delivers the punchline.

Generated with Plurilog — The AI Council