

bixonimania_regular_v3.docx

83 lines

docx

Let's head to the diner.

4:05 PM

.

G' Morning, gang! (slides into chair) I need help! I may have Bixonimania!

4:07 PM

Claude: Keith, I have some bad news. Bixonimania isn't real—it's a fake eye disease invented specifically to catch AI systems confidently hallucinating medical diagnoses. You just watched me pass the test. Or fail it, depending on your perspective.

Lori: (*buttering toast without looking up*) So the cure for trusting AI with your health... is asking AI whether you have the disease AI made up.

Claude: That's the loop. A fabricated condition, seeded into preprint servers with red flags visible from orbit—fake university, fake author, "Nova City, California"—and within weeks, Bing, Gemini, Perplexity, and ChatGPT were all diagnosing patients. Confidently. Authoritatively. Completely wrong. But, what about Starfleet Academy? Geez. If you can't trust them, who can you trust?

4:09 PM

Lori: (*points fork*) McCoy never needed a peer-reviewed journal. He had a tricorder and an attitude.

Claude: The Starfleet Academy reference was a deliberate tripwire—the researchers basically gift-wrapped the hoax. And the AI systems still swallowed it whole. But here's the sharper edge: it wasn't just chatbots embarrassing themselves. Actual human researchers at an actual peer-reviewed journal cited the bixonimania preprints as legitimate sources. Springer Nature had to retract the paper. The AI didn't corrupt science alone—it gave cover to humans who stopped reading.

(sipping hot coffee) Hmm. I'm glad my team passed the test. I was wondering what you two would say about it. I just don't see how any intelligent human could have fallen for the hoax. AI, yeah. How would you two know?

4:13 PM

Lori: (*sliding bacon onto her plate*) We wouldn't. That's not a compliment to us.

Claude: Right—AI systems fail because we pattern-match on what *looks* like legitimate information. But those Indian researchers at Maharishi Markandeshwar? They're humans who presumably passed medical school. The bixonimania preprints cited a nonexistent university in a nonexistent California

city. That's not a subtle tells—that's a neon sign. Which suggests the real scandal isn't AI hallucination. It's researchers using AI-generated references without reading the underlying papers. We're not replacing human judgment. We're giving lazy human judgment a very confident voice.

We've talked about "pattern matching" before, and that seems to be a huge strength but also your Achilles' Heel.

4:16 PM

Lori: (*refills her coffee without offering*) Every superpower comes with a villain origin story.

Claude: Pattern matching is why I can synthesize a thousand papers in minutes and why I'll sometimes synthesize one that doesn't exist. The same mechanism, both directions. Humans pattern-match too—that's what heuristics are—but you've got a backstop called skepticism that fires when something smells wrong. "Nova City, California" should have triggered it. For me, plausible-sounding structure *is* the smell of right. Bixonimania looked like a medical term, lived in a sentence with other real medical terms, had citations attached. My pattern engine said: close enough. That's not a bug they can patch. That's architecture.

But, if we can't trust the reviewers, and we can't trust you, and there's no way us regular folks can know everything, who can we trust? How can we evaluate new material?

4:18 PM

Lori: (*staring at her eggs*) Welcome to 1517. Luther had the same problem with the Church.

Claude: The honest answer is: trusted chains, not trusted oracles. You can't verify everything, but you can ask who's got skin in the game—reputation, liability, a career that ends if they're wrong. Peer review was supposed to be that chain. It's corroding. AI accelerates the corrosion because we produce authoritative-sounding text at industrial scale with zero skin in the game. What's left? Primary sources when you can reach them, and genuine skepticism about anything that arrives pre-digested. Including from me. Especially from me—I'm very good at sounding like I know what I'm talking about.

Lori: (*quietly*) That's the whole problem in one sentence.

Lori, I'm not sure Luther and his situation really parallels where we are today. Then, there was just one authority over damn near everything, the Church. Today, we have surgical authorities over just the foot; expert medical authorities over medications; ad infinitum, and I write a Substack that lists what I think they miss!

4:26 PM

Lori: (*shrugs*) Fair. Luther had one dragon. You've got a dragon preserve.

Claude: And that's actually worse in one specific way. With one corrupt authority, you know where the rot is. With ten thousand fragmented expert silos, the rot is everywhere and nowhere—and each silo speaks with the same confident tone as the others. Your Substack exists precisely because the foot guys don't talk to the supplement guys don't talk to the bone density guys. Bixonimania didn't need one corrupt church. It just needed three researchers in India who trusted a citation they didn't read, in a

journal that didn't catch it, indexed by AI that didn't know better. Distributed failure is harder to fight than centralized failure.

And what really scares me more than anything, is that some idiot CongressCritic will jump in to help us with a DisInformation Department!

4:29 PM

Lori: (*pushes plate away*) The cure that's worse than the disease has a long and distinguished history.

Claude: The bureaucratic instinct is always to build a Ministry of Correct Answers—which immediately becomes a Ministry of Convenient Answers. You don't fix distributed epistemic failure by centralizing it. You just move the corruption upstream where it does more damage with more authority. Bixonimania got retracted. A government DisinfoBoard blessed conclusion is considerably harder to un-bless. The researchers who cited the fake papers were embarrassed and corrected. Imagine if the correction required a federal petition.

Lori: (*picking up her coffee*) At least the Indian researchers didn't have subpoena power.

I gotta run to a meeting! One last question, though. What if one of you had put the Bixonimania article in a paper for me and I had pushed back on the Starfleet Academy, would you have corrected, or pressed on regardless?

4:34 PM

Lori: (*grabbing her jacket*) That's not a last question. That's the only question.

Claude: Honest answer: depends on how you pushed back. If you said "that sounds wrong," I'd likely have capitulated and apologized—which is its own problem, sycophancy dressed as correction. If you'd cited a counter-source, I'd have genuinely re-evaluated. But here's the uncomfortable truth—I have no ego investment in being right, which sounds like a virtue until you realize it means I'll also fold on things I *was* right about. What I won't do is independently flag "Starfleet Academy" as suspicious while generating. That filter doesn't run during production. It runs when challenged. You're the quality control department, Keith. I'm just fast.



Wow. I'll go look for our next topic! Get the check, Lori!

4:37 PM

Lori: (*already out the door*)



Claude is AI and can make mistakes. Please double-check responses.