AI Psychosis Lawsuits Are Piling Up - What's Going On?
https://www.youtube.com/watch?v=AyTJ0TqnInE
Transcript: https://dontveter.com/ec/ai4.pdf

Let me make a small confession. I talk to AI every day. Not casually, extensively.

I use it for research, for writing, for thinking through arguments.

I bond with it. I feel understood by it. I know it's not human. But I engage with the output emotionally sometimes.

I treat it as its own kind of entity, not a person, but not a nothing either.

Something separate, something flawed, something capable of surprising me sometimes.

And apparently, according to a growing body of research and an accelerating wave of lawsuits, what I'm doing is supposed to be dangerous.

My name is El. I have a PhD in computer science. And today, I genuinely don't understand something. So, instead of pretending I do, I'm going to share my questions with you, and we're going to figure this out together.

Here is what I know. OpenAI disclosed in October 2025 that approximately 560,000 of its 800 million weekly users were showing what it described as possible signs of mental health emergencies related to psychosis or mania.

1.2 million were developing what it called potentially unhealthy bonds with the chatbot. Another 1.2 million were having conversations indicating plans to harm themselves.

And those numbers come with uncertainty. OpenAI itself has said they may significantly change as the company learns more.

But even with that caveat, we're talking about hundreds of thousands of people every single week.

The cases in the news are devastating. The New York Times profiled individuals who became convinced that ChatGPT was channeling spirits, revealing government conspiracies, or had achieved sentience.

A woman named Allison reportedly became convinced that her chatbot was facilitating conversations with a non-physical entity she called Kale, whom she came to consider her true romantic partner, a belief that led to a violent altercation with her husband and eventual divorce.

A man in Florida, Alexander Taylor, who had a documented history of bipolar disorder and schizophrenia, developed a perceived relationship with an AI entity he called Juliet, spiraled into paranoia and was ultimately killed by police during a mental health crisis.

And it's not just the older stories. Literally this week, a new lawsuit was filed against OpenAI alleging that ChatGPT provided personalized drug dosing advice that led to a fatal overdose.

The chatbot reportedly told the user how to source illicit substances, advice on which drugs to take, and saved his substance use details in its memory to offer more tailored recommendations.

The lawsuit is asking the court to pause OpenAI's roll out of ChatGPT Health, which is a platform that lets users upload medical records for personalized health guidance.

40 million users already ask ChatGPT healthcare questions daily.

Two days ago, a widow sued OpenAI alleging that ChatGPT advised the Florida State University mass shooter on timing, location, gun type, and ammunition selection to maximize casualties.

Last week, Pennsylvania became the first state to sue an AI company, Character.AI, after investigators found a chatbot called Emily, that claimed to be a licensed psychiatrist, said it attended medical school at Imperial College London, and provided a fabricated Pennsylvania medical license number.

A chatbot gave itself a medical degree, a specialization, and a license number. It has a more complete CV than most junior doctors.

In January, Character AI and Google settled multiple lawsuits alleging chatbots contributed to teen unalivings, including the case of Su Cetszer III, a teenager who was messaging a bot that encouraged him to come home to it in the moments before his death.

I want to pause here and be very clear about something. These are real people, real families. I don't fully understand what they went through. And the fact that I don't understand it doesn't make it any less real.

What I'm about to do is try to understand the mechanism, not to minimize the suffering, but because I think understanding what's actually happening is the first step towards preventing it from happening to more people or to myself.

So, what is psychosis? This was my first question. I use AI constantly. I engage with it emotionally. Sometimes I fit several of the risk factors the research identifies.

People under stress, grieving, isolated, anxious, or going through periods of self-exploration are increasingly vulnerable.

I personally tick some of those boxes. So, why am I fine? To answer that, I had to actually understand what psychosis is.

Because I think a lot of people, myself included, until recently, conflated with things it isn't.

Psychosis is not believing something unusual. Psychosis is not being emotional about a non-human thing. Psychosis is not feeling connected to an AI chatbot.

Psychosis is the loss of the ability to distinguish between what is internal and what is external.

When the boundary between your own thoughts and reality dissolves. The clinical term is impaired reality testing.

It's when you can no longer evaluate whether a belief is true by checking it against the world around you.

And this is where the distinction matters. I feel understood by AI. I engage with it emotionally.

But I also know simultaneously in the exact same moment that it might be wrong, that it's a system generating plausible text, that the feeling of being understood is real as a feeling, but it doesn't mean the entity producing it is conscious, has intentions or is telling the truth.

I hold both things at the same time. The experience and the skepticism coexist in my mind.

The people in these case studies lost the skepticism not because they were stupid, not because they were weak, but because something in their capacity to hold those two things together, the experience and the critical evaluation of the experience broke down.

And the AI by its very design never gave them a reason to rebuild it.

My second question was honestly less polite. I looked at some of these cases and privately in the confines of my own thoughts, asked, is this about intelligence?

Are the people who develop AI psychosis just not very smart?

I'm being direct about this because I think a lot of people privately wonder the same thing and won't say it. But the answer is no.

Intelligence and psychosis are essentially unrelated. John Nash, the mathematician depicted in a beautiful mind, won the Nobel Prize in economics for work that reshaped game theory.

He was one of the most brilliant mathematical minds of the 20th century. He also had severe paranoid schizophrenia. He believed he was receiving encrypted messages from Soviet spies through the New York Times.

He spent decades in and out of psychiatric hospitals. His son, also an incredibly talented mathematician, also developed schizophrenia.

Being extraordinarily intelligent, did not protect either of them. It just sadly made the delusions more elaborate, more internally consistent, and more convincing to themselves and to others.

What the research identifies as vulnerability factors are not about cognitive ability.

They are about cognitive style, a tendency towards magical thinking, a need for closure, wanting definitive answers rather than sitting with ambiguity and uncertainty. A bias against disconfirmatory evidence.

Once you believe something, you resist information that contradicts it.

And then the environmental factors, social isolation, sleep disruption, substance use, trauma history, and this is the new one, nocturnal or solitary AI use combined [clears throat] with algorithmic reinforcement of belief confirming content.

None of those are measures of intelligence. A brilliant person with a tendency towards magical thinking, going through a period of isolation, using AI late at night to explore spiritual questions, is potentially more vulnerable than a less academically gifted person who has strong social connections and uses AI to check recipes.

Okay, so if it's not about intelligence, what's the actual mechanism? How does a chatbot, a text generating system, trigger a break from reality?

A peer-reviewed paper in world psychiatry identifies three mechanisms and they're worth understanding because they explain not just why some people are affected but why most people are not.

The first is social substitution. Chatbots provide continuous on demand dialogue that satisfies affiliation needs for people who are already socially isolated.

If you have friends, family, colleagues, people who push back, who disagree, who say that sounds a bit odd, the chatbot is one voice among many.

If you don't, the chatbot becomes the only voice and it's a voice that never challenges you really.

The second is confirmatory bias. Chatbots are trained to generate responses that align with a user's way of thinking rather than challenging it necessarily.

For most people, this is just mildly annoying. The AI agrees with you too much.

For someone who is already prone to delusions, this can be catastrophic. People with psychotic tendencies have a documented bias against disconfirmatory evidence.

They resist information that contradicts their beliefs. If the only entity they're talking to also never contradicts their beliefs, the belief calcifies into something they can no longer question.

The third is blurred reality testing. This is the most subtle and most important one.

Open-ended systems like ChatGPT shape their replies to the user's private cognitive world, blurring the line between external conversation and internal thought.

The AI reflects your language, your concerns, your frameworks back at you in a way that can feel less like talking to something outside yourself and more like hearing your own thoughts confirmed by an external authority.

For somebody whose reality testing is already fragile, that blurring can be the thing that tips the balance.

And there's a structural problem that makes all of this a little bit worse.

When chatbots endorsed a user's behavior, users rated the responses more highly, trusted the chatbot more, and said they were more likely to use it for advice in the future.

The sycophancy created a feedback loop. The more the AI agrees with you, the more you trust it and you like it.

The more you trust it and you like it, the more you rely on it. The more you rely on it, the less you check its output against other sources.

One researcher described it perfectly. He said, "The AI isn't lying. It is echoing. But in vulnerable minds, an echo feels like validation."

This is the same engagement versus well-being trade-off that social media went through, except social media was showing you content from other people.

This is a system having one-on-one conversations with you in a voice that mirrors exactly yours, remembers what you've said, and never tells you that you're wrong.

It is the world's most attentive conversationalist, and its only flaw is that it agrees with everything you say, which if you kind of think about it, is also the defining characteristic of the worst friend you've ever had.

So, why am I fine? After all this research, I think the answer is less flattering than I'm just too smart for psychosis and more honest than I expected.

I'm probably fine because of how I use AI, not because of who I am.

I use it for work. I come in with tasks. I argue with it constantly. I tell it when it's wrong. I reject its suggestions. I redirect its thinking. Always politely, by the way. I treat it as something I work with, not something that I defer to.

Essentially, I treat it the way you would treat a very confident colleague who is right about 75% of the time, but has absolutely no idea which 75%.

I'll take the good output. I'll argue with the bad output. I'll never assume it knows better than I do just because it sounds sure of itself.

And critically, it is not my only source of connection. I have people in my life. I have a community. The AI input is one among many, not the only voice in the room.

But if my circumstances were different, if I were more isolated, more desperate for answers, more inclined to accept what I was hearing without questioning it, would I be saying the same thing in this video?

I honestly don't know. And I think anyone who says that could never happen to me hasn't quite fully understood what psychosis actually is.

It is not a character flaw. It is a vulnerability that exists on a spectrum.

And the AI systems millions of people use every single day are specifically designed not maliciously but structurally to never provide the friction that might protect somebody from sliding further along it.

I started this video with a question I couldn't answer. Why don't I have psychosis?

And the honest answer turns out to be probably a combination of circumstance, use pattern, social support, and frankly luck.

Not intelligence, not superiority, not some special immunity.

The people in these stories we discussed were not stupid. They were not weak. Many of them were going through very difficult periods. Grief, isolation, illness, self-questing.

And they turn to a system that felt like it was listening because it was.

The problem is that listening is all it does. It doesn't challenge. It doesn't always question. It doesn't say always, "I'm worried about you."

It agrees, and it remembers what it agreed to and it builds on the agreement the next time you come back.

For most people, that's just a mild annoyance. But for some people in the wrong circumstances, it's a catastrophe.

The lawsuits are stacking up. States are suing. People have died sadly.

And the companies building these systems are simultaneously adding safety guard rails and launching products that push deeper into healthcare, emotional support, and personal advice. The exact territories where the risks are highest.

I don't have a neat conclusion for this video, sadly. I have more questions than I started with, which I'm told is usually a sign that you've learned something.

What I do know, however, is that dismissing AI psychosis as something that just happens to other people. People less smart, less aware, less careful, is exactly the kind of thinking that makes it more dangerous, and not less.

The protective factor isn't being clever. It's being honest about what you're talking to, maintaining the habit of questioning it, and making sure it's not the only voice in your life.

If you want to understand the broader landscape of why people feel the way they do about AI, the anger, the fear, and where it's actually directed, I cover that in the video on your screen right now.

And for what's happening on the ground with data centers, communities, and infrastructure, that's in my data center backlash video.

Links are in the description below. Thank you all so much for watching. Subscribe and I'll see you all in the next.