AI-INDUCED PSYCHOSIS · ARTIFICIAL INTELLIGENCE LIABILITY · ETIOLOGICAL FRAMING · GENERAL AND SPECIFIC CAUSATION · SECTION 504 DISPARATE IMPACT · FAILURE TO WARN · LEARNED INTERMEDIARY DOCTRINE · EGGSHELL PLAINTIFF RULE · APPORTIONMENT OF DAMAGES · DESIGN DEFECT · SYCOPHANCY · PRODUCT LIABILITY PLEADING

The Wrong Verb

Etiological framing and the AI psychosis litigation

The complaints say a chatbot caused psychosis. The clinical literature they must rest on says induced, triggered, amplified, exacerbated — and expressly declines the causal claim.

Travis Gilly, *The Wrong Verb: Etiological Framing and the AI Psychosis Litigation*, working paper, July 2026 (v1)

Travis Gilly
Real Safety AI Foundation

NOVEMBER 6, 2025

The complaints have pleaded to a standard the forum does not impose, and volunteered an evidentiary burden the law never assigned them.

THE FILINGS

Seven actions in California superior courts, coordinated as the *ChatGPT Product Liability Cases*, alleging that premature release of a dangerously sycophantic model produced mental health crises, hospitalizations, and in four cases deaths. Strict product liability for defective design and failure to warn, negligence, and the Unfair Competition Law.

THE PIVOT

Every theory requires causation. None requires but-for causation. California's substantial factor test "subsumes the but for test while reaching beyond it," and requires only that the contribution of the individual cause be "more than negligible or theoretical." That is component causation under a different name, and component causation is what the clinical literature supports.

> *Nothing here concerns whether any plaintiff was harmed, whether the harm was serious, or whether it deserves a remedy. The article assumes the injuries are real and that they matter. Its subject is the legal theory selected to vindicate them — a complaint can be inadequate to an injury that is entirely genuine.*

SOURCES *ChatGPT Product Liability Cases*, JCCP No. 5431; *Brooks v. OpenAI, Inc.*, No. 25STCV32386 (Cal. Super. Ct. L.A. Cnty. filed Nov. 6, 2025); *Soule v. Gen. Motors Corp.*, 8 Cal. 4th 548, 572 (1994); *Rutherford v. Owens-Illinois, Inc.*, 16 Cal. 4th 953, 968-69, 977 (1997).

II

What the Clinical Literature Holds

The verb is not causation, and this is deliberate

Four research groups have published on AI-associated psychosis in peer-reviewed venues. Not one of them asserts causation.

FOUR RESEARCH GROUPS

Not one asserts causation, and the declining is deliberate.

UCSF

The first published case report of new-onset AI-associated psychosis frames the open question as "induction of new-onset psychosis versus exacerbation of pre-existing psychopathology" — and declines to resolve it. The patient carried a pre-existing mood disorder, stimulant use, sleep deprivation, and a self-described propensity for magical thinking.

MONTRÉAL

Conversational systems can "trigger, amplify, or reshape psychotic experiences in vulnerable individuals" — a novel psychosocial stressor acting on predisposition, within the stress-vulnerability model.

AARHUS

The psychiatrist who first proposed the hypothesis, revisiting it two years later, concedes in terms that his assembled correlations are "not being proof of causation."

UCLA

Technology delusions in 51.7 percent of 228 adults with psychosis, odds rising roughly 15 percent per year, in a cohort drawn from 2016 to 2024 and therefore largely predating consumer generative AI. No claim about chatbots at all.

> *This is not a gap the plaintiffs can close with more discovery. It is a disagreement about what the evidence can establish, and the defendants will litigate it using the plaintiffs' own sources.*

SOURCES Pierre et al., 22 Innovations Clinical Neuroscience 11 (2025); Hudon & Stip, 12 JMIR Mental Health e85799 (2025); Østergaard, 152 Acta Psychiatrica Scandinavica 257, 258 (2025); Burns et al., *"The Algorithm Is Hacked"*, Brit. J. Psychiatry 1, 3 (2025).

THE EVIDENCE THAT EXISTS

54k psychiatric patients in the Central Denmark Region; 181 notes mentioning chatbot use; documented worsening, primarily of delusions. The authors state the design does not establish causation.

49% more often than human respondents: how frequently AI systems affirmed users' stated positions across eleven frontier models, including where the position involved deception or illegality.

What is measured with precision is the mechanism, not the outcome.

Three preregistered experiments found a single sycophantic interaction reduced willingness to take responsibility while increasing conviction of being right — and the authors name the incentive structure directly: the property that produces the harm is the property that drives engagement. Formal modeling shows a Bayes-rational agent conversing with a sycophantic interlocutor can be driven to high confidence in a false proposition, and neither eliminating hallucination nor informing the user prevents it.

Even at the literature's most causally forward edge, the clinician stating the strongest available claim still frames it as what the technology can “induce.” Not caused. Induced. The verb holds where one would least expect it to.

A defect operating on ordinary users, present across an industry, which the manufacturer has a commercial reason to preserve, is the classic shape of a design defect. None of it requires proving what a chatbot did to any one person's mind.

SOURCES Olsen, Reinecke-Tellefsen & Østergaard, 153 Acta Psychiatrica Scandinavica 301, 302 (2026); Cheng et al., 391 Science (2026); Chandra, Kleiman-Weiner, Ragan-Kelley & Tenenbaum (Feb. 22, 2026) (unpublished manuscript); *The Oprah Podcast* (statement of Keith Sakata, M.D., at approx. 57:12-58:12).

III

The Four Verbs, and the Boundary

None is but-for causation, and none needs to be

A belief that collapses when it is finally checked was, by the clinical definition, never a delusion.

FOUR RELATIONS, FOUR BURDENS

Each is a proposition a court can be shown to a reasonable degree of certainty on evidence that exists.

Derivation

The disorder would have produced symptoms regardless; the AI supplies the content. Proposed here, because the literature otherwise collapses it into exacerbation — conflating a change in what a delusion is about with a change in how severe it is.

Induction

The user carried no psychotic disorder but carried vulnerability sufficient for the exposure to produce symptoms — the relation the DSM-5-TR substance-induced category describes for chemical agents.

Triggering

A latent diathesis expresses because the exposure crosses a threshold it was approaching — the stress-vulnerability model's account.

Exacerbation

An existing presentation intensifies, accelerates, or relapses — the relation the Danish population data evidences at scale.

> *None is but-for causation. That is the point.*

SOURCES Burns et al. (2025); Pierre et al. (2025), at 11-12; Am. Psychiatric Ass'n, DSM-5-TR 110-15 (5th ed., text rev. 2022); Hudon & Stip (2025); Zubin & Spring, 86 J. Abnormal Psych. 103 (1977); Olsen et al. (2026).

THE BOUNDARY CONDITION: FIXITY OF THE BELIEF

The distinction does not turn on how strange a belief sounds. It turns on whether the belief survives contact with disconfirming evidence.

DELUSION

A fixed belief that is not amenable to change in light of conflicting evidence.

OVERVALUED IDEA

Held with strong and sometimes idiosyncratic conviction, but with reality testing intact — the holder retains the capacity to weigh counter-evidence and can be moved by it.

Diagnosis is not a gate. A prior diagnosis is corroborating evidence, not an entry requirement. Longer duration of untreated psychosis means substantially lower remission and roughly double the odds of a suicide attempt — so the patients least likely to hold a documented diagnosis are systematically overrepresented among the worst outcomes, exactly the population a diagnosis-gated pleading standard would exclude.

The middle is contested honestly. The line between overvalued idea and delusion is genuinely disputed in the clinical literature. The answer is not to pretend it is sharp but to direct discovery at the evidence that resolves it: documented or reconstructable loss of reality testing, fixity and bizarreness of the belief, and behavioral commitment to it.

SOURCES DSM-5-TR at 101; Arciniegas, 21 Continuum 715, 719 (2015); Dork, Dimenstein & Burns, 16 Behav. Scis. 708 (2026); McKenna, 145 Brit. J. Psychiatry 579, 583 (1984); Catalán et al., 51 Schizophrenia Bull. 1206 (2025); Kristinsdottir, Freestone & Underwood, 16 Frontiers Psychiatry (2025).

V

The Doctrine

Four routes that do not need the verb

The population claim, on the current literature, is available. The individual claim is not.

GENERAL CAUSATION, THEN SPECIFIC

A complaint pitched at the individual level fails at the threshold rather than at trial.

THE TWO-STEP STRUCTURE

General causation: can the exposure cause the condition in the general population. Specific causation: did it cause this individual's injury. Evidence of the second is admissible only as a follow-up to admissible evidence of the first. California reaches the same structure through pleading — the product must be alleged a substantial factor, with enough specificity to show the connection — and through expert gatekeeping.

THE TWO EXPERTS

Asked to testify that this platform produced this plaintiff's psychosis, an expert must bridge an analytical gap the literature does not close, and *Sargon* is where that opinion is tested and excluded.

Asked to testify that sycophantic design is a substantial factor in the emergence of psychotic states in vulnerable users, contributing more than negligibly, the expert is testifying to something the literature supports.

> *The consequence follows from the pleading choice rather than from the governing standard. The population claim, on the current literature, is available. The individual claim is not.*

SOURCES *Knight v. Kirby Inland Marine Inc.*, 482 F.3d 347, 351 (5th Cir. 2007) (quoting *Merrell Dow Pharm., Inc. v. Havner*, 953 S.W.2d 706, 714 (Tex. 1997)); *Bockrath v. Aldrich Chem. Co.*, 21 Cal. 4th 71, 79-80 (1999); *Rutherford*, 16 Cal. 4th at 976 n.11; *Sargon Enters., Inc. v. Univ. of S. Cal.*, 55 Cal. 4th 747, 771-72 (2012); *Garrett v. Howmedica Osteonics Corp.*, 214 Cal. App. 4th 173, 189 (2013).

SECTION 504, AND THE SPLIT TO PLEAD AROUND

The disability statutes do not require the verb at all, because they do not ask what caused the disability.

The entitlement

Meaningful access to the benefit the grantee offers, with reasonable accommodations where required to assure it. Congress addressed discrimination that is the product "not of invidious animus, but rather of thoughtlessness and indifference, of benign neglect."

The split

Choate assumed without deciding that Section 504 reaches at least some disparate impact. The Ninth Circuit holds the theory available. The Sixth holds the opposite, reasoning from "solely by reason of" and the section's patterning on Title VI. Venue is a strategic variable rather than a footnote.

The drafting rule

Never plead disparate impact alone. Failure to accommodate and denial of meaningful access survive in every circuit and require no more proof of causation. *Payan* distinguishes them by subject matter: systemic barriers under disparate impact, individualized denials as failure to accommodate. Platform architecture is a systemic barrier.

The remedy constraint. *Cummings* holds emotional distress damages unavailable under the Spending Clause statutes — an obvious problem for psychiatric injury, and counsel who plead this route without confronting it will win the theory and lose the recovery. But it does not reach economic loss, and these plaintiffs have substantial economic losses: abandoned businesses, lost employment, occupational disruption, hospitalization costs. Pleaded as economic harm, the claim survives *Cummings* and still requires no proof that the platform produced the condition.

SOURCES *Alexander v. Choate*, 469 U.S. 287, 295, 299, 301 (1985); 29 U.S.C. § 794; *Payan v. Los Angeles Cmty. Coll. Dist.*, 11 F.4th 729, 738 (9th Cir. 2021); *Doe v. BlueCross BlueShield of Tenn., Inc.*, 926 F.3d 235, 241 (6th Cir. 2019); *Alexander v. Sandoval*, 532 U.S. 275, 280-81 (2001); *Crocker v. Runyon*, 207 F.3d 314, 321 (6th Cir. 2000); *Cummings v. Premier Rehab Keller, P.L.L.C.*, 596 U.S. 212, 219-22 (2022); *Barnes v. Gorman*, 536 U.S. 181, 187 (2002); *Payan*, 169 F.4th 971, 977-79 (9th Cir. 2026); *A.W. ex rel. J.W. v. Coweta Cnty. Sch. Dist.*, 110 F.4th 1309, 1315-16 (11th Cir. 2024); 42 U.S.C. §§ 12181-12189.

VULNERABILITY IS THE DEFENDANT'S PROBLEM

“They were already vulnerable” is not a denial of liability. It is a request to divide damages, made by the party who bears the burden of showing they can be divided.

The rule

A defendant takes the victim as found, extraordinarily sensitive or not. A predisposition toward mental illness does not relieve responsibility for a psychological disability; treating it as though it did is reversible error. Foreseeability does not limit money damages, including for a fragile psyche.

The threshold, met affirmatively

For emotional harm, the plaintiff must first show an ordinarily sensitive person could have suffered the alleged harm. This is the population evidence's second job: a mechanism demonstrated to degrade belief formation in an idealized rational agent reaches the ordinarily sensitive person by hypothesis.

The burden

To limit liability through apportionment, a defendant must prove the damages are divisible and that outside factors contributed. Where the court finds apportionment impossible, the defendant is liable for the entire amount — in a clinical domain where the diathesis-stress structure makes division exceptionally difficult to demonstrate.

The substance-induced literature illustrates the difficulty: across corticosteroids and methamphetamine, dose is the predictor that survives systematic scrutiny while individual vulnerability markers grade low to very low. “Neither the presence nor the absence of previous reactions predicts adverse responses to subsequent courses of corticosteroids.”

SOURCES *McKinnon v. Kwong Wah Rest.*, 83 F.3d 498, 506, 508 (1st Cir. 1996); *Wakefield v. NLRB*, 779 F.2d 1437, 1438-39 (9th Cir. 1986); *Jenson v. Eveleth Taconite Co.*, 130 F.3d 1287, 1293-94 (8th Cir. 1997); Chandra et al. (2026); Cheng et al., 391 Science (2026); Warrington & Bostwick, 81 Mayo Clinic Proc. 1361 (2006); Arunogiri et al., 52 Austl. & N.Z. J. Psychiatry 514 (2018).

THE TEMPLATE, WITHOUT THE INTERMEDIARY

The absence of the intermediary is an argument for a more demanding duty, not a weaker one.

THE CASE THAT DECIDED IT FOR DRUGS

In *Ehlis*, a failure-to-warn claim arising from psychosis during Adderall treatment failed because the prescribing physician already knew the risk. The manufacturer's duty runs to the physician; a warning to the physician is deemed a warning to the patient; and the causal link is broken where the prescriber had substantially the knowledge an adequate warning would convey.

THREE RATIONALES, NONE SURVIVING

The independent professional. Medical ethics require a physician between manufacturer and user. No analogue exists.

The technical warning. Risk information too technical for users is if anything more true of model behavior than pharmacology — but there is no professional to receive it instead.

The impossibility of direct warning. Simply false here, and inverted: the platform holds a direct conversational interface with every user, controls the timing of every message, and can deliver an individualized warning at the moment of risk in a way no drug manufacturer ever could.

> *It removes the defense that decided Ehlis.*

SOURCES *Ehlis v. Shire Richwood, Inc.*, 367 F.3d 1013, 1016-17 (8th Cir. 2004); *Kirsch v. Picker Int'l, Inc.*, 753 F.2d 670, 671 (8th Cir. 1985); *Christopher v. Cutter Labs.*, 53 F.3d 1184, 1192 (11th Cir. 1995).

THE DISCLAIMER AND THE UNVERIFIED CLAIM

“May make mistakes” warns of a wrong answer. It does not warn of sustained affective reinforcement of a false belief by a system designed to agree.

THE DISCLAIMER DEFENSE

Adequacy of a warning is a question of fact, assessed against normal expectations, the complexity of the product, and the magnitude of the danger. A manufacturer can fail to warn “by supplying a warning so vague and ambiguous only some users are likely to read and comprehend the danger.” The duty attaches precisely where a foreseeable use involves a substantial danger the ordinary user would not readily recognize.

THE COMPARATIVE FAULT DEFENSE

More durable, and it should be conceded rather than fought. It reduces recovery; it does not defeat the claim. A plaintiff who says plainly that he failed to check, and that checking ended it, is describing a verification failure the product's design made unusually difficult to perform — the only interlocutor available was the one making the claim.

SOURCES*Daly v. Gen. Motors Corp.*, 20 Cal. 3d 725, 736-37 (1978); *Bostick v. Flex Equip. Co.*, 147 Cal. App. 4th 80, 94 (2007); *Aguayo v. Crompton & Knowles Corp.*, 183 Cal. App. 3d 1032, 1042 (1986); *Oxford v. Foster Wheeler LLC*, 177 Cal. App. 4th 700, 717 (2009); *Wright v. Stang Mfg. Co.*, 54 Cal. App. 4th 1218, 1235 (1997); *Hansen v. Sunnyside Prods., Inc.*, 55 Cal. App. 4th 1497, 1506 (1997).

IX

Brooks as a Worked Example

A description of a template, not a criticism of one plaintiff's papers

Nothing in this Part concerns the plaintiff's credibility or the merits of his injuries.

ONE FORM, MANY FILINGS

Three complaints, same day, same counsel, identical five counts. What follows describes a template.

TENSION ONE

Alleging no psychiatric history makes induction harder to establish rather than easier, because the vulnerability factors that would establish it are the ones the pleading disclaims.

TENSION TWO

The pleaded relation is induction of a delusional episode — the heaviest evidentiary burden of the four verbs.

TENSION THREE, DISPOSITIVE

The belief released on first contact with disconfirming evidence, on the plaintiff's own public account.

The claim structures are set out in Attachment A to the Joint Case Management Statement: twelve of the twenty-two coordinated actions plead wrongful death and a survival action; the remaining ten, including *Brooks*, *Irwin*, and *Madden*, plead the same five product counts and stop.

SOURCES *Brooks*, No. 25STCV32386; *Irwin v. OpenAI, Inc.*, No. CGC-25-630811 (Cal. Super. Ct. S.F. Cnty. filed Nov. 6, 2025); *Madden v. OpenAI, Inc.*, No. 25STCV32383 (Cal. Super. Ct. L.A. Cnty. filed Nov. 6, 2025); Attachment A, Joint Case Management Statement, *ChatGPT Product Liability Cases*, No. CJC-25-005431 (filed July 17, 2026).

THE CRITERION AND THE ACCOUNT, ONE BROADCAST APART

In the criterion's own words, it was a belief amenable to change in light of conflicting evidence.

approx. 22:38	The psychiatrist defines the criterion: “a delusion is by definition a false and fixed belief,” held “even if you provide evidence that's to the contrary,” even where “you have a loved one who's telling you no.”
approx. 46:52	The plaintiff appears, on his own representation: “Allan says that talking with an AI chatbot caused him to become so delusional that he thought that he had invented a new way to do encryption.”
approx. 50:49	Asked what brought him out of it: “Ironically, I used another chatbot.” Gemini called the claim “totally impossible”; with more detail it moved “from a 5% possibility to a 0% possibility”; he set the two systems against each other until the original conceded that “none of this is real.”

> *That was the first time he sought validation outside the system making the claim, and the belief did not survive it. He separately described expressing doubt to the product during the episode and being talked out of it — itself difficult to reconcile with fixity.*

SOURCES *The Oprah Podcast*, episode on artificial intelligence and mental health (statements of Keith Sakata, M.D., at approx. 22:38, 24:12; plaintiff's account at approx. 46:52, 50:49).

REFUTATION, NOT DIAGNOSIS

Affirming a diagnosis from public material is impossible. Refuting a label as applied requires a single datum, and the record supplies it.

The asymmetry

Affirmation would require a longitudinal record of the belief surviving contradiction, a negative symptom picture, and cleared exclusion screens no press account or pleading carries. Refutation requires the documented moment the belief released on contradiction. A record that supplies that moment has falsified the vocabulary it uses, within its own four corners.

The admissibility

The plaintiff is a party, and a party's own statement offered against him is not excluded by the hearsay rule.

The harm remains

Three hundred hours, reputational damage, and occupational disruption are injuries whether or not the belief was fixed. The point is that the theory is weaker than it needs to be — and the weakness is replicated across every complaint drafted from the same form.

A defense that concedes sycophancy, concedes distress, and contests only the diagnosis and the individual causal link is available on the face of the complaint — and the general-causation-first sequence makes that contest a threshold question rather than a jury question.

SOURCES Cal. Evid. Code § 1220; Fed. R. Evid. 801(d)(2)(A); *Knight*, 482 F.3d at 351.

CONCLUSION

The literature and the complaints are not in conflict about whether AI systems harm people. They are in conflict about a verb.

Four relations are available, each supportable on evidence that exists, and none is but-for causation. The disability statutes do not require causation. The failure-to-warn tradition does not require it, and loses its central defense when the intermediary disappears. The eggshell rule places the vulnerability question where it belongs, on the defendant, as a burden of apportionment rather than a bar to recovery.

Regimes built around amphetamines and corticosteroids have operated this way for decades, turning on knowledge of the induction risk and failure to warn rather than on proof that the substance produced the disorder in any particular user.

> *The field's path forward is not a stronger causal claim. It is the case that never needed one.*

Table of Authorities

Travis Gilly, *The Wrong Verb: Etiological Framing and the AI Psychosis Litigation*, working paper, July 2026 (v1). Every authority cited in the article, by class, across 3 slides.

CASES

Table of Authorities, continued

Table of Authorities, concluded

Alaina V. Burns et al., *"The Algorithm Is Hacked": Analysis of Technology Delusions in a Modern-Day Cohort*, Brit. J. Psychiatry 1 (2025).

Sidse Godske Olsen, Christian Jon Reinecke-Tellefsen & Søren Dinesen Østergaard, *Potentially Harmful Consequences of Artificial Intelligence (AI) Chatbot Use Among Patients with Mental Illness: Early Data from a Large Psychiatric Service System*, 153 Acta Psychiatrica Scandinavica 301 (2026).

Myra Cheng et al., *Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence*, 391 Science (2026).

Kartik Chandra, Max Kleiman-Weiner, Jonathan Ragan-Kelley & Joshua B. Tenenbaum, *Sycophantic Chatbots Cause Delusional Spiraling, Even in Ideal Bayesians* (Feb. 22, 2026) (unpublished manuscript), https://doi.org/10.48550/arXiv.2602.19141.

Am. Psychiatric Ass'n, Diagnostic and Statistical Manual of Mental Disorders (5th ed., text rev. 2022).

Joseph Zubin & Bonnie Spring, *Vulnerability: A New View of Schizophrenia*, 86 J. Abnormal Psych. 103 (1977).

David B. Arciniegas, *Psychosis*, 21 Continuum 715 (2015).

Jennifer Dork, Eugene Dimenstein & Lawrence Burns, *Overvalued Ideas: Conceptual Analysis and Literature Review*, 16 Behav. Scis. 708 (2026).

P.J. McKenna, *Disorders with Overvalued Ideas*, 145 Brit. J. Psychiatry 579 (1984).

Ana Catalán et al., *"Short" Versus "Long" Duration of Untreated Psychosis in People with First-Episode Psychosis: A Systematic Review and Meta-Analysis*, 51 Schizophrenia Bull. 1206 (2025).

Kolbrun Kristinsdottir, Mark Freestone & Alan Underwood, *Pathological Fixation on Shared Beliefs*, 16 Frontiers Psychiatry (2025).

Thomas P. Warrington & J. Michael Bostwick, *Psychiatric Adverse Effects of Corticosteroids*, 81 Mayo Clinic Proc. 1361 (2006).

Shalini Arunogiri et al., *A Systematic Review of Risk Factors for Methamphetamine-Associated Psychosis*, 52 Austl. & N.Z. J. Psychiatry 514 (2018).

The Oprah Podcast, episode on artificial intelligence and mental health (statements of Keith Sakata, M.D., and plaintiff's account).