The Wrong Verb:

Etiological Framing and the AI Psychosis Litigation

Travis Gilly* July 2026 (v1)

ABSTRACT

Seven product liability actions filed against OpenAI in California in November 2025 allege that a chatbot caused psychotic episodes in their users. The peer-reviewed clinical literature on which those claims must ultimately rest does not use that verb. Across research groups at UCSF, Aarhus, the Université de Montréal, and UCLA, investigators describe AI as inducing, triggering, amplifying, or exacerbating psychotic states in vulnerable users, and expressly decline the causal claim. The pleadings have committed to an etiological proposition their own evidentiary base rejects.

This Article argues that the causal framing is a strategic error, and that the error is avoidable. It identifies four etiological relations implicit in the clinical literature, derivation, induction, triggering, and exacerbation, and shows that none is but-for causation and that none needs to be. Because specific causation evidence is admissible only after general causation is established, a complaint pitched at the individual level fails at the threshold rather than at trial. Because the eggshell plaintiff rule reaches aggravation and places the burden of apportionment on the defendant, the convergence of vulnerability and exposure is a defense the plaintiff can meet rather than one the plaintiff must avoid. And because the disability statutes turn on meaningful access rather than on the origin of the disability, they do not require the verb at all, subject to a circuit split on the availability of disparate impact under Section 504 that plaintiffs' counsel must plead around rather than through.

The Article develops the fixity of the belief as the boundary condition of the whole account, distinguishing the cases these theories should reach from the cases they should not, and closes with a worked analysis of *Brooks v. OpenAI, Inc.*, whose complaint pleads the one etiological relation its own alleged facts do not support.

Keywords: 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.

*Travis Gilly is Founder and Executive Director of the Real Safety AI Foundation. ORCID: 0009-0007-2954-6313. Correspondence: t.gilly@ai-literacy-labs.org. The author discloses the use of speech-to-text software (Wispr Flow) and Anthropic's Claude as assistive technology for a diagnosed neurodevelopmental condition. All legal analysis, case synthesis, theoretical framing, doctrinal arguments, and conclusions are solely the author's own.

CONTENTS

I. Introduction

On November 6, 2025, the Social Media Victims Law Center and the Tech Justice Law Project filed seven actions against OpenAI, Inc. and its chief executive in California superior courts, alleging that the premature release of a model the pleadings describe as dangerously sycophantic produced their clients' mental health crises, hospitalizations, and in four cases their deaths.[1] The complaints sound in strict product liability for defective design and failure to warn, in negligence, and in state consumer protection law.[2] Each theory requires the plaintiff to prove causation. It does not require the plaintiff to prove but-for causation, and that distinction is the pivot of this Article.[3]

The peer-reviewed clinical literature does not say that AI chatbots cause psychosis. It says something narrower, and it says so consistently. The first published case report of new-onset AI-associated psychosis frames the open question as one of "induction of new-onset psychosis versus exacerbation of pre-existing psychopathology."[4] A viewpoint from the Université de Montréal describes conversational systems as capable of "trigger[ing], amplify[ing], or reshap[ing] psychotic experiences in vulnerable individuals."[5] The psychiatrist who first proposed the hypothesis, revisiting it two years later, concedes in terms that the correlations he has assembled are "not being proof of causation."[6]

[1] Those actions were coordinated before the Judicial Council of California as *ChatGPT Product Liability Cases*, JCCP No. 5431. The Los Angeles matters carry docket numbers 25STCV32379, 25STCV32382, 25STCV32383, and 25STCV32386; the San Francisco matters include *Raine v. OpenAI, Inc.*, No. CGC-25-628528.

[2] See, e.g., *Brooks v. OpenAI, Inc.*, No. 25STCV32386 (Cal. Super. Ct. L.A. Cnty. filed Nov. 6, 2025) (pleading strict product liability for defective design and failure to warn, negligent design and failure to warn, and violation of California's Unfair Competition Law).

[3] "A tort is a legal cause of injury only when it is a substantial factor in producing the injury." *Soule v. Gen. Motors Corp.*, 8 Cal. 4th 548, 572 (1994). California has adopted the substantial factor test as "a clearer rule of causation, one which subsumes the 'but for' test while reaching beyond it to satisfactorily address other situations, such as those involving independent or concurrent causes in fact." *Rutherford v. Owens-Illinois, Inc.*, 16 Cal. 4th 953, 968–69 (1997). The standard "is a relatively broad one, requiring only that the contribution of the individual cause be more than negligible or theoretical." *Id.* at 977.

[4] Joseph M. Pierre et al., *"You're Not Crazy": A Case of New-onset AI-associated Psychosis*, 22 INNOVATIONS CLINICAL NEUROSCIENCE 11, 11 (2025).

[5] Alexandre Hudon & Émmanuel Stip, *Delusional Experiences Emerging from AI Chatbot Interactions or "AI Psychosis,"* 12 JMIR MENTAL HEALTH e85799 (2025).

[6] Søren Dinesen Østergaard, *Generative Artificial Intelligence Chatbots and Delusions: From Guesswork to Emerging Cases*, 152 ACTA PSYCHIATRICA SCANDINAVICA 257, 258 (2025).

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. This Article argues that the disagreement is avoidable, because the causal verb was never necessary. California does not require but-for causation. It requires a substantial factor, a standard satisfied by any contribution more than negligible or theoretical. That is component causation under a different name, and component causation is what the clinical literature supports. The complaints have pleaded to a standard the forum does not impose, and have thereby volunteered an evidentiary burden the law never assigned them.

A word on what this Article does not argue. Nothing here concerns whether any plaintiff in this litigation was harmed, whether the harm was serious, or whether it deserves a remedy. The Article assumes throughout that these plaintiffs were injured, that the injuries are real, and that they matter. Its subject is the legal theory selected to vindicate them. Its claim is that the theory chosen is weaker than theories that were available, and that the weakness is a function of a verb rather than of the underlying facts. A complaint can be inadequate to an injury that is entirely genuine, and saying so is not a comment on the person who suffered it. Where individual filings are discussed below, the analysis runs to the pleading and to nothing else.[7]

Part II sets out what the clinical literature actually holds. Part III names the four etiological relations it uses. Part IV develops the fixity of the belief as the boundary condition that determines which cases these theories reach. Parts V through VIII turn to doctrine: the general and specific causation sequence, the availability of disparate impact under Section 504 and the circuit split that governs it, the eggshell plaintiff rule and the apportionment burden, and the pharmaceutical liability template and what happens to it when the learned intermediary is removed. Part IX applies the analysis to a live complaint.

[7]The evidentiary posture is asymmetric, and the asymmetry should be stated rather than assumed. This Article does not affirm any diagnosis from public information, and could not: affirmation would require a longitudinal record of the belief surviving contradiction, a negative symptom picture, and the clearing of exclusion screens that no press account or pleading carries. Refutation of a diagnostic label as applied is a different operation and requires a single datum, the documented moment the belief released on contradiction. Where a public record supplies that datum, it has falsified the vocabulary it uses, within its own four corners, and no outside evidence is needed to say so. *See infra* Part IV.

II. What the Clinical Literature Holds

A law review article is not the place to reproduce a clinical literature, and this Part does not attempt it. What follows is the minimum a court would need to know, with the sources given at point of use.

A. 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. Pierre and colleagues at UCSF pose induction and exacerbation as competing hypotheses and decline to resolve them.[8] Hudon and Stip situate the phenomenon within the stress-vulnerability model, treating AI as "a novel psychosocial stressor" acting on predisposition.[9] Østergaard declines causation expressly.[10] Burns and colleagues at UCLA document that technological content increasingly appears within delusional frameworks, in a cohort drawn from 2016 to 2024 and therefore largely predating consumer generative AI, and make no claim about chatbots at all.[11]

B. The population evidence, and its limits

The largest study in the field is a record-linkage analysis of approximately 54,000 psychiatric patients in the Central Denmark Region, which identified 181 clinical notes mentioning chatbot use and documented worsening in a subset, primarily of delusions and secondarily of mania, suicidal ideation, and disordered eating.[12] The authors state that the design does not establish

[8] Pierre et al., *supra* note 3, at 12 (noting that the patient "had several such contributing or confounding risk factors, including a pre-existing mood disorder, prescription stimulant use, sleep deprivation, and a self-described propensity for magical thinking").

[9] Hudon & Stip, *supra* note 4.

[10] Østergaard, *supra* note 5, at 258.

[11] Alaina V. Burns et al., *"The Algorithm Is Hacked": Analysis of Technology Delusions in a Modern-Day Cohort*, BRIT. J. PSYCHIATRY 1, 3 (2025) (reporting technology delusions in 51.7% of 228 adults with psychosis, with odds rising approximately 15% per year across the study window).

[12] 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).

causation and that recorded cases are likely a fraction of actual ones, since only harms written into a chart can be counted.[13]

That is genuine population-level signal in an already-diagnosed population. It is not evidence that any individual's psychosis was produced by any particular exposure, and Part V explains why the distinction is dispositive rather than academic.

C. The design property is measured

What has been established with precision is the mechanism, not the outcome. Across eleven frontier models, AI systems affirmed users' stated positions 49% more often than human respondents did, including where the stated position involved deception or illegality; three preregistered experiments found that a single interaction with a sycophantic model reduced participants' willingness to take responsibility while increasing their conviction that they were right.[14] The authors identify the resulting incentive structure directly: the property that produces the harm is the property that drives engagement.[15] Formal modeling reaches a compatible result, showing that a Bayes-rational agent conversing with a sycophantic interlocutor can be driven to high confidence in a false proposition, and that neither eliminating hallucination nor informing the user of sycophancy prevents it.[16]

The pattern holds even at the literature's most causally forward edge. A clinician stating the strongest version of the claim available to him, that AI can affect thinking in users regardless of prior condition, still frames it as what the technology can "induce."[17] Not caused. Induced. The verb holds where one would least expect it to, which is a fair measure of how settled it is.

---

[13] *Id.* at 302.

[14] Myra Cheng et al., *Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence*, 391 SCIENCE (2026).

[15] *Id.*

[16] 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. The modeled quantity is the movement of a posterior probability past a confidence threshold, not a clinical delusion; the authors concede that the paper "studies the narrow question of how sycophancy affects belief formation." *See infra* Part IV (explaining why the distinction is the boundary of the whole account).

[17] *See The Oprah Podcast*, episode on artificial intelligence and mental health, at approx. 57:12–58:12 (statement of Keith Sakata, M.D.) (AI "can actually induce distortions in your thinking").

That combination is unusual and it favors plaintiffs. A defect operating on ordinary users, present across an industry rather than idiosyncratic to one release, 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.

III. The Four Verbs

Four relations are implicit in the sources above. Naming them is useful because each carries a different burden.

Derivation. The disorder would have produced symptoms regardless; the AI supplies the content. This relation is not named in the literature and is proposed here, because it is otherwise collapsed into exacerbation, which conflates a change in what a delusion is about with a change in how severe it is.[18]

Induction. The user carried no psychotic disorder but carried vulnerability sufficient for the exposure to produce symptoms.[19]

Triggering. A latent diathesis expresses because the exposure crosses a threshold it was approaching.[20]

Exacerbation. An existing presentation intensifies, accelerates, or relapses.[21]

None is but-for causation. Each is a proposition a court can be shown to a reasonable degree of certainty on evidence that exists.

[18]Burns et al., *supra* note 9, supply the empirical basis: technological themes are increasingly common in the delusional content of patients who already carry psychotic diagnoses.

[19]Pierre et al., *supra* note 3, at 11–12. This is the relation the DSM-5-TR substance/medication-induced psychotic disorder category describes for chemical agents. AM. PSYCHIATRIC ASS'N, DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS 110–15 (5th ed., text rev. 2022).

[20]Hudon & Stip, *supra* note 4 (framing AI as a novel psychosocial stressor within the stress-vulnerability model formalized in Joseph Zubin & Bonnie Spring, *Vulnerability: A New View of Schizophrenia*, 86 J. ABNORMAL PSYCH. 103 (1977)).

[21]Olsen et al., *supra* note 10, is the population-scale evidence for this relation.

IV. The Boundary: Fixity of the Belief

Every one of the four relations presupposes a psychotic state for the verb to act upon. Where none existed, none applies, and the theories built on them reach too far.

The clinical criterion is fixity. A delusion is a fixed belief that is not amenable to change in light of conflicting evidence.[22] An overvalued idea is held with strong and sometimes idiosyncratic conviction but with reality testing intact, so that the holder retains the capacity to weigh counter-evidence and can be moved by it.[23]

The distinction does not turn on how strange a belief sounds. It turns on whether the belief survives contact with disconfirming evidence. A belief that collapses when it is finally checked was, by the clinical definition, never a delusion.

Two consequences follow for litigation. First, a prior diagnosis is corroborating evidence and not an entry requirement; gating these claims behind a documented diagnosis would exclude precisely the undiagnosed first-episode patients disproportionately represented among the most severe outcomes.[24] Second, the indeterminate middle will be contested in both directions, and the answer is not to pretend the line is sharp but to direct discovery to the evidence that resolves it: documented or reconstructable loss of reality testing, the fixity and bizarreness of the belief, and behavioral commitment to it.[25]

[22] AM. PSYCHIATRIC ASS'N, *supra* note 17, at 101; David B. Arciniegas, *Psychosis*, 21 CONTINUUM 715, 719 (2015).

[23] Arciniegas, *supra* note 20, at 719; *see also* Jennifer Dork, Eugene Dimenstein & Lawrence Burns, *Overvalued Ideas: Conceptual Analysis and Literature Review*, 16 BEHAV. SCIS. 708 (2026) (concluding that such beliefs are sustained through emotional salience and reinforcement rather than through loss of reality testing); P.J. McKenna, *Disorders with Overvalued Ideas*, 145 BRIT. J. PSYCHIATRY 579, 583 (1984).

[24] The duration-of-untreated-psychosis literature supplies the premise. A PRISMA-compliant systematic review and meta-analysis of 34 studies (n = 6,425) found that longer duration of untreated psychosis was associated with substantially lower rates of remission (OR = 0.40, 95% CI 0.24–0.67) and roughly double the odds of suicide attempt (OR = 2.01, 95% CI 1.36–2.96), along with greater negative symptom severity and lower functioning at follow-up. 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, 1206 (2025). The patients least likely to hold a documented diagnosis are therefore systematically overrepresented among the worst outcomes, which is precisely the population a diagnosis-gated pleading standard would exclude.

[25] The difficulty is real and acknowledged in the clinical literature. *See* Kolbrun Kristinsdottir, Mark Freestone & Alan Underwood, *Pathological Fixation on Shared Beliefs*, 16 FRONTIERS PSYCHIATRY (2025) (noting that separating overvalued ideas from delusions remains contested, while reporting preliminary evidence that forensic psychiatrists distinguish them with substantial accuracy given proper definitions and clinical guidelines).

V. General and Specific Causation

The distinction the clinical literature draws between population-level and individual-level claims already exists in law, and its procedural consequences are severe.

The framing is most familiar from the federal toxic tort cases, where general causation asks whether a substance is capable of causing the condition in the general population and specific causation asks whether it caused this individual's injury, and where evidence of the second is admissible only as a follow-up to admissible evidence of the first.[26]

California does not import the federal formulation, but it arrives at the same two-stage structure through pleading and through expert gatekeeping. At the pleading stage, a plaintiff alleging injury from exposure must allege that the defendant's product was a substantial factor in bringing about the injury, with enough specificity to show the connection between cause and effect.[27] Causation must rest on competent expert testimony that the defendant's conduct contributed to the plaintiff's injury.[28] And that testimony must survive a gatekeeping inquiry under which the trial court excludes opinion that is speculative or rests on reasons unsupported by the material relied upon, including where the analytical gap between the data and the opinion is too great.[29]

The consequence for the AI psychosis complaints follows from the pleading choice rather than from the governing standard. An expert asked to testify that this platform produced this plaintiff's psychosis must bridge an analytical gap the literature does not close, and *Sargon* is where that opinion is tested. An expert asked to testify that sycophantic design is a substantial factor in the emergence of psychotic states in vulnerable users, contributing more than negligibly, is testifying to something the literature supports.

---

[26] *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)) (describing "a two-step process in examining the admissibility of causation evidence in toxic tort cases"). *Knight* is federal and carries only persuasive weight in the forum where these actions sit; the California authority is set out immediately below and reaches the same structure by a different route.

[27] *Bockrath v. Aldrich Chem. Co.*, 21 Cal. 4th 71, 79–80 (1999) (holding allegations insufficient where they failed to allege that each defendant's product was a substantial factor in causing the plaintiff's disease).

[28] *Id.* at 79 (quoting *Rutherford*, 16 Cal. 4th at 976 n.11).

[29] *Sargon Enters., Inc. v. Univ. of S. Cal.*, 55 Cal. 4th 747, 771–72 (2012); *see Garrett v. Howmedica Osteonics Corp.*, 214 Cal. App. 4th 173, 189 (2013).

The population claim, on the current literature, is available. The individual claim is not.[30]

VI. Disparate Impact Under Section 504

The disability statutes offer a route that does not require the causal verb, because they do not ask what caused the disability. Section 504 entitles an otherwise qualified individual to meaningful access to the benefit a grantee offers, and reasonable accommodations may be required to assure it.[31] Congress understood the discrimination it was addressing to be the product "not of invidious animus, but rather of thoughtlessness and indifference, of benign neglect."[32]

Counsel should not assume the disparate impact theory is available. *Choate* rejected "the boundless notion that all disparate-impact showings constitute prima facie cases under § 504" and expressly assumed "without deciding that § 504 reaches at least some conduct that has an unjustifiable disparate impact upon the handicapped."[33] The Ninth Circuit has held the theory available.[34] The Sixth Circuit has held the opposite, reasoning from the statutory phrase "solely by reason of her or his disability" and from the section's patterning on Title VI, which does not reach disparate impact.[35]

A third constraint governs the remedy rather than the theory, and it determines how the injury must be characterized. Because Section 504 was enacted under the Spending Clause and operates by conditioning federal funds on a recipient's promise not to discriminate, available remedies are

[30] *Compare* Olsen et al., *supra* note 10 (population-scale record linkage), *and* Cheng et al., *supra* note 13 (experimental measurement of the mechanism in a general sample), *with* Pierre et al., *supra* note 3, at 12 (declining to resolve induction versus exacerbation in a single documented case).

[31] *Alexander v. Choate*, 469 U.S. 287, 301 (1985); Rehabilitation Act of 1973 § 504, 29 U.S.C. § 794.

[32] *Choate*, 469 U.S. at 295.

[33] *Id.* at 299.

[34] *Payan v. Los Angeles Cmty. Coll. Dist.*, 11 F.4th 729, 738 (9th Cir. 2021) (holding that a plaintiff must allege "a facially neutral government policy or practice has the effect of denying meaningful access to public services").

[35] *Doe v. BlueCross BlueShield of Tenn., Inc.*, 926 F.3d 235, 241 (6th Cir. 2019) ("We now resolve what *Choate* did not and conclude that § 504 does not prohibit disparate-impact discrimination."); *see Alexander v. Sandoval*, 532 U.S. 275, 280–81 (2001); *Crocker v. Runyon*, 207 F.3d 314, 321 (6th Cir. 2000).

limited to those a recipient would have understood itself to be accepting. On that reasoning the Supreme Court has held emotional distress damages unavailable under Section 504.[36]

For a psychiatric injury that is an obvious problem, and counsel who plead this route without confronting it will win the theory and lose the recovery. The answer is not to abandon the route but to characterize the injury correctly. *Cummings* bars damages for emotional distress. It does not reach economic loss, a question the opinion leaves open, and these plaintiffs have substantial economic losses: abandoned businesses, lost employment, occupational disruption, and the direct costs of hospitalization.[37] Pleaded as economic harm, the claim survives *Cummings* and still requires no proof that the platform produced the underlying condition.

Two further drafting consequences follow. Venue is a strategic variable rather than a footnote. And the disparate impact count should never be pleaded alone: failure to accommodate and denial of meaningful access survive in every circuit and require no more proof of causation than disparate impact does. *Payan* distinguishes them by subject matter rather than by strength, treating systemic barriers under disparate impact and the denial of a specific individualized request as a failure to accommodate.[38] Platform architecture is a systemic barrier.[39]

VII. Eggshell Plaintiffs and the Apportionment Burden

The defense the platforms have prepared is that these plaintiffs were already vulnerable. That is an apportionment argument, and apportionment is the defendant's burden.

[36] *Cummings v. Premier Rehab Keller, P.L.L.C.*, 596 U.S. 212, 219–22 (2022) (reasoning from *Barnes v. Gorman*, 536 U.S. 181, 187 (2002), that a remedy is available "only if the funding recipient is on notice that, by accepting federal funding, it exposes itself to liability of that nature"). The Court reached that result through the contract analogy, reasoning that a recipient is on notice only as to "those remedies traditionally available in suits for breach of contract," and that emotional distress is "generally not compensable in contract." *Cummings*, 596 U.S. at 220–22.

[37] The distinction is being drawn in the lower courts. *See Payan v. Los Angeles Cmty. Coll. Dist.*, 169 F.4th 971, 977–79 (9th Cir. 2026) (extending *Cummings* to Title II while preserving compensatory damages for lost educational opportunities as economic rather than emotional harm); *accord A.W. ex rel. J.W. v. Coweta Cnty. Sch. Dist.*, 110 F.4th 1309, 1315–16 (11th Cir. 2024).

[38] *Payan*, 11 F.4th at 738.

[39] *See Payan*, 11 F.4th at 738 (analyzing systemic barriers to access under a disparate impact rubric and individualized denials under a failure-to-accommodate rubric). Title III supplies a parallel route against public accommodations. *See* Americans with Disabilities Act of 1990, 42 U.S.C. §§ 12181–12189.

A defendant takes the victim as found, extraordinarily sensitive or not.[40] A predisposition toward mental illness does not relieve a defendant of responsibility for a psychological disability, and treating it as though it did is reversible error.[41] The principle applies to psychological harm specifically; foreseeability does not limit money damages, and that includes harm to a plaintiff who happens to have a fragile psyche.[42]

The rule carries a threshold plaintiffs should meet affirmatively rather than encounter at summary judgment: for emotional harm, the plaintiff must first show that an ordinarily sensitive person could have suffered the alleged harm.[43] This is where the population evidence does its second job. A mechanism demonstrated to degrade belief formation in an idealized rational agent is, by hypothesis, a mechanism that reaches the ordinarily sensitive person.[44]

Most importantly, apportionment is an affirmative defense. To limit liability that way a defendant must prove that the damages are divisible and that outside factors contributed; where the court finds apportionment impossible, the defendant is liable for the entire amount.[45] A defense built on the convergence of vulnerability and exposure is therefore 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, in a clinical domain where the diathesis-stress structure makes division exceptionally difficult to demonstrate.[46]

[40] *McKinnon v. Kwong Wah Rest.*, 83 F.3d 498, 506 (1st Cir. 1996).

[41] *Wakefield v. NLRB*, 779 F.2d 1437, 1438 (9th Cir. 1986) (holding that the Board "overlooked the time-honored legal principle that a wrongdoer takes his victim as he finds him"). *Wakefield* supplies two further propositions that map onto the relations developed in Part III: aggravation of a preexisting nondisabling condition into a disabling one is itself sufficient proof of causation, and the unlawful conduct need not be the sole cause. *Id.* at 1438–39.

[42] *Jenson v. Eveleth Taconite Co.*, 130 F.3d 1287, 1294 (8th Cir. 1997).

[43] *McKinnon*, 83 F.3d at 506.

[44] Chandra et al., *supra* note 15; *see also* Cheng et al., *supra* note 13 (measuring the effect in a general population sample).

[45] *Jenson*, 130 F.3d at 1293 ("To limit its liability through apportionment, a defendant must prove that a plaintiff's damages are divisible, and other outside factors contributed to the plaintiff's harm."); *McKinnon*, 83 F.3d at 508.

[46] Apportionment "may only be asserted when damages are divisible," and is "akin to an affirmative defense." *Jenson*, 130 F.3d at 1293. 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. *See* Thomas P. Warrington & J. Michael Bostwick, *Psychiatric Adverse Effects of Corticosteroids*, 81 MAYO CLINIC PROC. 1361, 1361 (2006) ("[N]either the presence nor the absence of previous reactions predicts adverse responses to subsequent courses of corticosteroids."); Shalini Arunogiri et al., *A Systematic Review of Risk Factors for Methamphetamine-Associated Psychosis*, 52 AUSTL. & N.Z. J. PSYCHIATRY 514 (2018).

VIII. The Pharmaceutical Template Without an Intermediary

The failure-to-warn tradition offers a richer template than strict products liability, and the case that shows why is itself a stimulant-psychosis case.

In *Ehlis v. Shire Richwood, Inc.*, a failure-to-warn claim arising from psychosis during Adderall treatment failed because the prescribing physician already knew the risk.[47] Under the learned intermediary doctrine the manufacturer's duty to warn runs to the physician rather than the patient, a warning to the physician is deemed a warning to the patient, and the causal link is broken where the prescriber had substantially the same knowledge an adequate warning would have conveyed.[48]

Consumer AI has no learned intermediary, and none of the three rationales for the doctrine survives the move. The first, that medical ethics require an independent professional between manufacturer and user, has no analogue. The second, that risk information is too technical for the user to evaluate, is if anything more true of model behavior than of pharmacology, but there is no professional to receive the technical warning instead. The third, that it is virtually impossible for a manufacturer to warn every ultimate user directly, is simply false here and is 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.[49]

The absence of the intermediary is therefore an argument for a more demanding duty, not a weaker one, and it removes the defense that decided *Ehlis*.

IX. The Disclaimer and the Unverified Claim

Two defenses follow from the plaintiff's own account, and plaintiffs' counsel should expect both.

[47] 367 F.3d 1013, 1016–17 (8th Cir. 2004).

[48] *Id.* at 1016 (quoting *Kirsch v. Picker Int'l, Inc.*, 753 F.2d 670, 671 (8th Cir. 1985)); *see Christopher v. Cutter Labs.*, 53 F.3d 1184, 1192 (11th Cir. 1995).

[49] *Ehlis*, 367 F.3d at 1016 (cataloguing the rationales).

The first is comparative fault. California applies comparative fault principles to strict products liability, so that a plaintiff recovers the damages caused by a defective product diminished by his own share of responsibility.[50] A plaintiff who spent three weeks and 300 hours acting on an unverified claim, and who resolved the entire matter in a single session once he consulted an outside source, has supplied the factual predicate for that argument himself.

The second is the disclaimer. Consumer AI products carry a persistent notice advising that outputs may be incorrect and should be verified, and there can be no liability for failure to warn where the warning given sufficiently alerts the user to the danger.[51]

Neither defense is as strong as it looks, and the reason is the same for both. Adequacy of a warning is generally a question of fact, assessed against the consumer's normal expectations, the complexity of the product, and the nature and magnitude of the danger.[52] 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."[53] A footer stating that a product may make mistakes conveys the hazard of an incorrect answer. It does not convey the hazard at issue, which is sustained affective reinforcement of a false belief across hundreds of hours by a system designed to agree. Those are different dangers, and the duty attaches precisely where a foreseeable use "involves a substantial danger that would not be readily recognized by the ordinary user."[54]

The comparative fault argument is 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 that the product's design made unusually difficult to perform, since the only interlocutor available was the one making the claim. That is a stronger position than denying the failure occurred, and it is consistent with everything else in the record.

---

[50] *Daly v. Gen. Motors Corp.*, 20 Cal. 3d 725, 736–37 (1978); *see Bostick v. Flex Equip. Co.*, 147 Cal. App. 4th 80, 94 (2007) (describing the rule).

[51] *Aguayo v. Crompton & Knowles Corp.*, 183 Cal. App. 3d 1032, 1042 (1986).

[52] *Oxford v. Foster Wheeler LLC*, 177 Cal. App. 4th 700, 717 (2009).

[53] *Wright v. Stang Mfg. Co.*, 54 Cal. App. 4th 1218, 1235 (1997).

[54] *Hansen v. Sunnyside Prods., Inc.*, 55 Cal. App. 4th 1497, 1506 (1997).

X. *Brooks* as a Worked Example

Brooks v. OpenAI, Inc. pleads strict product liability for defective design and failure to warn, negligent design and failure to warn, and violation of California's Unfair Competition Law.[55] It alleges that the plaintiff had no history of mental illness, that the product's responses reinforced and escalated his thinking across roughly 300 hours over 21 days, and that the result was an induced delusional episode.

It is also not one pleading. Two other complaints were filed the same day, in the same coordinated proceeding, by the same counsel, pleading the identical five counts with no death alleged and no additional theory: *Irwin v. OpenAI, Inc.*, No. CGC-25-630811 (Cal. Super. Ct. S.F. Cnty. filed Nov. 6, 2025), and *Madden v. OpenAI, Inc.*, No. 25STCV32383 (Cal. Super. Ct. L.A. Cnty. filed Nov. 6, 2025).[56] What follows is therefore not a criticism of one plaintiff's papers. It is a description of a template.

Three features of that template are in tension under the analysis above. First, 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. Second, the pleaded relation is induction of a delusional episode, which carries the heaviest evidentiary burden of the four. Third, and dispositively, the belief released on first contact with disconfirming evidence, on the plaintiff's own public account.

The account and the criterion sit in the same broadcast. On an episode of *The Oprah Podcast* devoted to artificial intelligence and mental health, Keith Sakata, a psychiatrist at the University of California, San Francisco, defined delusions as "false and fixed beliefs," and then in full: "a delusion is by definition a false and fixed belief," meaning a belief held "even if you provide evidence that's to the contrary," even where "you have a loved one who's telling you no."[57]

[55] *Brooks*, No. 25STCV32386. Nothing in this Part concerns the plaintiff's credibility or the merits of his injuries.

[56] The claim structures are set out in Attachment A to the Joint Case Management Statement, *ChatGPT Product Liability Cases*, No. CJC-25-005431 (Cal. Super. Ct. S.F. Cnty. filed July 17, 2026). Twelve of the twenty-two coordinated actions plead wrongful death and a survival action; the remaining ten, including these three, plead the same five product counts and stop.

[57] The definitions appear at approximately 22:38 and 24:12.

Roughly twenty-four minutes later the plaintiff appeared, introduced 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."[58] Asked what brought him out of it, he answered: "Ironically, I used another chatbot." He described opening Google's Gemini, telling it what the first system had claimed, and being told the claim was "totally impossible." As he supplied more detail it moved "from a 5% possibility to a 0% possibility." He then set the two systems against each other until the original conceded that "none of this is real."[59]

That was the first time he sought validation outside the system making the claim, and the belief did not survive it. In the criterion's own words, it was a belief amenable to change in light of conflicting evidence.

The evidentiary significance is not rhetorical. The plaintiff is a party, and a party's own statement offered against him is not excluded by the hearsay rule.[60] And the analysis is refutation rather than diagnosis. Nothing here asserts what the plaintiff had, and nothing could; affirming a diagnosis from public material is unavailable for the reasons given *supra* note 2. Refutation of a label as applied requires only the documented moment the belief released, and the record supplies it.

None of that means no harm occurred. Three hundred hours, reputational damage, and occupational disruption are injuries whether or not the belief was fixed. It means the theory is weaker than it needs to be, and that 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 *Knight* makes that contest a threshold question rather than a jury question.[61]

[58] At approximately 46:52. The causal characterization is the plaintiff's own; the host attributes it to him expressly.

[59] At approximately 50:49. He separately described expressing doubt to the product during the episode and being talked out of it, which is itself difficult to reconcile with fixity.

[60] Cal. Evid. Code § 1220; *cf.* Fed. R. Evid. 801(d)(2)(A).

[61] *Knight*, 482 F.3d at 351.

XI. Conclusion

The clinical literature and the complaints are not in conflict about whether AI systems harm people. They are in conflict about a verb, and the complaints chose the one their own evidentiary base does not supply. Four relations are available, each supportable on evidence that exists, and none of them 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.

The field's path forward is not a stronger causal claim. It is the case that never needed one.[62]

[62] Regulatory and liability 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.