AI-INDUCED PSYCHOSIS · FIXITY · DELUSION · OVERVALUED IDEA · DSM-5-TR · DIAGNOSTIC LANGUAGE · DILUTION · CONCEPT CREEP

The Fixity Test

Refutation without diagnosis in the AI-induced psychosis discourse

Journalism, advocacy, and litigation attach the words "delusion" and "psychosis" to named individuals from public information. This commentary argues the practice fails on the diagnostic manual's own terms, and that the failure runs in one direction. Confirming the label from a public record is impossible in principle. Refuting it takes exactly one datum the record routinely supplies: the documented moment the belief released on contradiction.

Travis Gilly · Real Safety AI Foundation

Gilly, T. (2026). The Fixity Test: Refutation Without Diagnosis in the AI-Induced Psychosis Discourse. Working paper, July 2026 (v0.4).

The people asserting the diagnosis most confidently are the people least positioned to assert it.

Reporters, advocates, and pleadings deploy "delusion" and "psychosis" with the manual's full authority, converting **an account of a hard episode into an account of a medical event.**

The clinicians publishing in peer review do the opposite: Pierre, Østergaard, Hudon & Stip, Keshavan all **decline to diagnose the named individuals, and hedge the category itself.** The looseness lives downstream of the clinic.

Sources Pierre (2025); Østergaard (2023, 2025); Hudon & Stip (2025); Keshavan et al. (2026); Gilly, sec. I.

PART I OF III

The Test: Fixity

The definition turns on one structural question, and it is a property of trajectory

Gilly, T. (2026). The Fixity Test. Working paper, July 2026 (v0.4), sec. II.

> ***A delusion is "a fixed belief that is not amenable to change in light of conflicting evidence." When contradiction arrives, does the belief hold?***

Not how bizarre it sounds, how completely it dominates a life, or how alarmed the family is. Those are content and intensity, the properties the definition sets aside.

The **overvalued idea** can be just as intense and destructive, but it retains reality testing, and yields to counter-evidence. Even its forensic extreme is still, explicitly, not a delusion.

Sources American Psychiatric Association, DSM-5-TR (2022); McKenna (1984); Rahman (2018); Gilly, sec. II.

A photograph of a ball in midair cannot say whether it is rising or falling.

A snapshot of a belief at its peak cannot say whether it is fixed. Fixity lives only in **trajectory**, in the belief's encounters with contradiction over time.

RETROSPECTIVE, AND DECISIVE ONE WAY

A belief that released on sustained contradiction **was never a delusion.** It did not stop being one at the moment of release, it failed fixity, and the failure reaches backward through the whole episode. **The word never applied.**

Sources American Psychiatric Association, DSM-5-TR (2022); Gilly, sec. II.

PART II OF III

The Asymmetry

Confirmation is impossible from a snapshot. Refutation needs one datum the record supplies.

Gilly, T. (2026). The Fixity Test. Working paper, July 2026 (v0.4), secs. III–IV.

TO AFFIRM "DELUSION" — UNAVAILABLE

Needs a trajectory of the belief **surviving** contradiction; "psychosis" needs the negative-symptom cluster too, absences invisible to the genre. And two exclusion screens, culture and schizotypy, run before fixity is even reached. **Missing in principle from a snapshot.**

TO REFUTE IT — ONE DATUM

The documented moment the belief **released on contradiction.** Fixity failed there, the requirement was never met, and the word is refuted by the record itself. The psychosis claim rests on the delusion claim, so both fall together.

The observer claims no clinical authority. They are reading a definition and a document, and reporting that the document fails the definition it invokes.

Sources American Psychiatric Association, DSM-5-TR (2022); Eckblad & Chapman (1983); Gilly, sec. III.

The same genres that misuse the words are built to write the refutation into the document.

THE JOURNALISM RECOVERY ARC

A feature needs an ending, so it opens with "delusion," peaks, and closes with the belief dissolving, often in one conversation. **The closing beat is the fixity test failed on the page.**

THE LITIGATION DAMAGES NARRATIVE

A complaint recites induced delusion, then the plaintiff's return to reality, the return is what lets them sue. **The plaintiff filed the falsifier.**

Not every record self-refutes, some show the belief holding for months, through hospitalization, to death. Those it marks as where the words **might** belong. The test discriminates.

Refutation without diagnosis.

MAY DO

Describe the content, trajectory, and harm. Classify the pattern the record displays. Demonstrate what the record **rules out**, a released belief matches the overvalued idea and fails the definition of delusion.

MAY NEVER DO

Assert what the individual **has.** Refuting one label licenses no substitute label. The person may have any condition or none, that belongs to a clinician with access to them, and it stays open.

It says what the record cannot support. It does not say, and cannot say, what is true of the person.

Sources Gilly, sec. V, Refutation Without Diagnosis.

Refusing the psychiatric costume does not shrink the injury. It names it correctly.

Careers disrupted, families strained, months lost inside a belief a system amplified for engagement, that is a **consumer injury with its own legal vocabulary and remedies.** Harm does not need the word "psychosis" to matter.

TWO PILES, AND NO MORE

Public information sorts records into **refuted** and **unrefuted.** Whether a surviving record is an acute episode or a spectrum condition, everything finer, is clinical work, not the public's to perform.

PART III OF III

The Cost of the Borrowed Word

The expense lands on two populations at once, from the same act

Gilly, T. (2026). The Fixity Test. Working paper, July 2026 (v0.4), sec. VI.

One description points at the product. The other points at the person's brain.

POPULATION 1 · THE MISLABELED

Handed a psychiatric identity, "I was psychotic", in place of a consumer injury, carrying into employment, custody, insurance, self-concept. **The word relocates the defect from the product to the user.**

POPULATION 2 · THE DIAGNOSED

The community whose beliefs did not release and whose families measure the illness in years. For them "psychosis" is **load-bearing, and its usefulness depends on meaning something specific.**

Sources Gilly, sec. VI, The Cost of the Borrowed Word.

"Concept creep," running at headline speed, and the harm is documented, not asserted.

28%

of newspaper articles mentioning schizophrenia used it as a metaphor, against 1% for cancer (Duckworth et al.).

Label

alone worsened attitudes toward schizophrenia specifically, with near-zero effect for depression (Angermeyer & Matschinger).

2 dec.

of population studies: acceptance did not improve even as biological literacy rose (Schomerus et al.).

Disease trivialization, the casual application of a diagnostic term to benign behavior, does measurable work on audiences. **Every borrowed "psychosis" repeats the operation at higher stakes.**

Sources Rüsch et al. (2005); Angermeyer & Matschinger (2003); Schomerus et al. (2012); Pavelko & Myrick (2015, 2016, 2017); Duckworth et al. (2003); Yeung et al. (2022); Haslam (2016); Gilly, sec. VI.

A mislabeled case is built to lose on the word, and each loss is a citation against the cases that carry the true signature.

Guarding the boundary of the word is not gatekeeping sympathy. **It is protecting the precedent, the diagnosis, and the community the word exists to serve.**

The definition was written to exclude the culturally shared and the yielding, and reserve the category for the fixed. The discourse expands it onto exactly what it was written to exclude. **Not a kindness to anyone.**

Sources Haslam (2016); American Psychiatric Association, DSM-5-TR (2022); Gilly, sec. VI.

The definition supplies the test, fixity. The genres supply the data, the recovery arc and the damages narrative, each writing the moment of release into the same document that asserts the word. Where the release is documented, the belief was never a delusion, the episode was never psychosis, and the record has said so itself.

The discourse need not choose between taking the harm seriously and using the words correctly. Using the words correctly is what taking the harm seriously looks like.

Sources Gilly, The Fixity Test, sec. VIII, Conclusion.

References

All references as listed in Gilly, T. (2026), The Fixity Test, Working paper, July 2026 (v0.4). Reproduced verbatim (1 of 2).

American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). American Psychiatric Publishing.

Angermeyer, M. C., & Matschinger, H. (2003). The stigma of mental illness: Effects of labelling on public attitudes towards people with mental disorder. Acta Psychiatrica Scandinavica, 108(4), 304–309.

Duckworth, K., Halpern, J. H., Schutt, R. K., & Gillespie, C. (2003). Use of schizophrenia as a metaphor in U.S. newspapers. Psychiatric Services, 54(10), 1402–1404.

Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. Journal of Consulting and Clinical Psychology, 51(2), 215–225.

Haslam, N. (2016). Concept creep: Psychology's expanding concepts of harm and pathology. Psychological Inquiry, 27(1), 1–17.

Hudon, A., & Stip, E. (2025). Delusional experiences emerging from AI chatbot interactions or "AI psychosis." JMIR Mental Health, 12, e85799.

Keshavan, M., Torous, J., & Yassin, W. (2026). Do generative AI chatbots increase psychosis risk? World Psychiatry, 25(1), 150–151.

McKenna, P. J. (1984). Disorders with overvalued ideas. British Journal of Psychiatry, 145(6), 579–585.

Østergaard, S. D. (2023). Will generative artificial intelligence chatbots generate delusions in individuals prone to psychosis? Schizophrenia Bulletin, 49(6), 1418–1419.

Østergaard, S. D. (2025). Generative artificial intelligence chatbots and delusions: From guesswork to emerging cases. Acta Psychiatrica Scandinavica, 152(4), 257–259.

Pavelko, R. L., & Myrick, J. G. (2015). That's so OCD: The effects of disease trivialization via social media on user perceptions and impression formation. Computers in Human Behavior, 49, 251–258.

Pavelko, R. L., & Myrick, J. G. (2016). Tweeting and trivializing: How the trivialization of obsessive–compulsive disorder via social media impacts user perceptions, emotions, and behaviors. Imagination, Cognition and Personality, 36(1), 41–63.

Pavelko, R. L., & Myrick, J. G. (2017). Examining differences in audience recall and reaction between mediated portrayals of mental illness as trivializing versus stigmatizing. Journal of Health Communication, 22(11), 876–884.

Pierre, J. M. (2025). Can AI chatbots validate delusional thinking? BMJ, 391, r2229.

References (continued)

Reproduced verbatim (2 of 2).

Rahman, T. (2018). Extreme overvalued beliefs: How violent extremist beliefs become "normalized." Behavioral Sciences, 8(1), 10.

Rüsch, N., Angermeyer, M. C., & Corrigan, P. W. (2005). Mental illness stigma: Concepts, consequences, and initiatives to reduce stigma. European Psychiatry, 20(8), 529–539.

Schomerus, G., Schwahn, C., Holzinger, A., Corrigan, P. W., Grabe, H. J., Carta, M. G., & Angermeyer, M. C. (2012). Evolution of public attitudes about mental illness: A systematic review and meta-analysis. Acta Psychiatrica Scandinavica, 125(6), 440–452.

Yeung, A., Ng, E., & Abi-Jaoude, E. (2022). TikTok and attention-deficit/hyperactivity disorder: A cross-sectional study of social media content quality. The Canadian Journal of Psychiatry, 67(12), 899–906.

THE TEST, IN ONE LINE

Confirmation is impossible from the snapshot. Refutation is available wherever the record contains the release.