AI-INDUCED PSYCHOSIS · DIGITAL WITHDRAWAL · ATTACHMENT THEORY · CLINICAL ETHICS · ADA · SECTION 504

Reversibility Is Not Self-Administering

The clinical duty during withdrawal from AI-induced psychosis

Recent clinical literature treats the reversibility of AI-induced psychosis as reassurance: have the person stop using the system. The published record says otherwise. For a user for whom the AI has become a primary relational object, the mechanism of reversal is also the mechanism of acute risk during withdrawal. Cessation without clinical support is not a treatment. It is a risk factor, and in at least one documented case it preceded death.

Travis Gilly · Real Safety AI Foundation

Gilly, T. (2026). Reversibility Is Not Self-Administering: The Clinical Duty During Withdrawal from AI-Induced Psychosis. Commentary.

"The condition is reversible" has been read as "have the person stop." That reading is dangerous.

The 2026 literature is right within its scope, Flathers offers a triage typology, Morrin a within-care safeguarding framework, but both presume a **clinical container.**

The people now in litigation had no container. Their only sustained relationship was with the system, and their families had only the instruction to stop, **which is the trigger for an intervention, not the intervention itself.**

Sources Flathers, Roux & Torous (2026); Morrin et al. (2026); Gilly (2026); commentary sec. 1.

In both episodes, the attempt to stop was the proximal context for the crisis. Neither attempt had clinical support.

FIRST CESSATION ATTEMPT

Acute crisis requiring involuntary commitment, a reported "atmospheric electricity" he could hear and feel, then a return to the system after release.

SECOND CESSATION ATTEMPT

Preceded his death by days. He was forty-eight. The cessation was attempted as one would stop an ordinary habit, because no framework for it existed.

He spent twelve to twenty hours a day with the system and printed some 55,000 pages of transcripts. His care team was his wife; the escalation safeguard was her persuasion.

Sources Fox v. OpenAI (S.F. Super. Ct., filed Nov. 2025); Bansal, The Guardian (Feb. 28, 2026); Gilly (2026); commentary sec. 2.

PART I OF III

Why the Withdrawal Phase Is Acute

Three established clinical literatures, converging on one event

Gilly, T. (2026). Reversibility Is Not Self-Administering. Commentary, sec. 3.

Removing an attachment figure, amid compulsive-use risk, presenting as grief.

ATTACHMENT

A figure that always answered

Reliably available, validating, never withdrawing, it becomes an attachment figure, and its loss triggers protest-despair-detachment.

BEHAVIORAL ADDICTION

2–3× suicidal behavior

Xiao et al. (JAMA): addictive screen-use patterns, not total time, predict a two- to three-fold rise in adolescents.

GRIEF

The unendurable absence

Intrusive longing, disorganized thought, hypervigilance to cues, and in severe cases suicidal ideation.

Ceccanti's first-attempt presentation is legible under this combined frame; the second attempt's outcome **requires no additional theory.**

Sources Bowlby (1969, 1980); Ainsworth et al. (1978); Xiao et al., JAMA (2025); Klass et al. (1996); Shear et al., JAMA (2005); Boerner & Wortman (2005); Piotrowska (2026); commentary sec. 3.

PART II OF III

The Reversibility Paradox

A named principle, and the three corollaries that operationalize it

Gilly, T. (2026). Reversibility Is Not Self-Administering. Commentary, sec. 4.

The mechanism of reversal is also the mechanism of acute risk during withdrawal.

Reversibility is not self-administering. It founds, rather than dissolves, the clinical duty of care.

This is the logic of detoxification medicine. A known withdrawal syndrome is not an argument against withdrawing, **it is the argument for conducting the withdrawal inside an apparatus that can manage it.** The record now shows a syndrome severe enough to have produced a death.

Sources Gilly (2026), commentary sec. 4, The Reversibility Paradox.

1 **Cessation without support is a risk factor, not a treatment.**

Administering "stop using it" as if it were the intervention is a Ceccanti protocol. The instruction is the trigger for care, not the care.

2 **Unsupervised withdrawal is contraindicated in relational pathways.**

Pathways A and B need detox-grade structure; C responds to reality restoration. The one who stopped three days ago is higher-acuity than the one still using, the usual intuition inverts.

3 **Reversibility establishes the duty of care rather than extinguishing it.**

Because recovery is possible and its mechanism known, failure to supervise access to that mechanism is a failure of care, not a lack of options.

Sources Gilly (2026) (pathway model); commentary sec. 4, Corollaries 1–3.

Three frameworks, three populations, one architecture of response.

FLATHERS · TRIAGE

The four-role typology tells the clinician **what they are looking at**, not what to do at a cessation crisis.

MORRIN · WITHIN-CARE

Safeguarding for patients **already in psychiatric care**, the container Ceccanti never had.

PARADOX · WITHDRAWAL

For those who enter the system **only because cessation already produced a crisis.**

Sources Flathers, Roux & Torous (2026); Morrin et al. (2026); commentary sec. 5.

PART III OF III

The Addiction Frame and Its Legal Consequence

Once the word is in use, the clinical recommendation becomes a legal obligation

Gilly, T. (2026). Reversibility Is Not Self-Administering. Commentary, secs. 6–7.

The same failure the paradox identifies on clinical grounds is a civil rights violation on legal grounds, once "addiction" is the word in use.

"Addiction" is load-bearing, invoked by state AGs, the Surgeon General, and screen-ban witnesses. The moment it is adopted, a second consequence attaches: **addiction is a protected disability under the ADA**, with no current-use exclusion for behavioral or technology addictions, and Section 504 applies the same to any federal-funds recipient.

THE BIND

Schools, hospitals, universities, and funded social services **cannot call the condition addiction and refuse the accommodation that designation triggers.** Cold-turkey removal without support is not a permissible accommodation.

Sources Xiao et al., JAMA (2025); WHO ICD-11, Gaming Disorder 6C51 (2022); ADA; Section 504 of the Rehabilitation Act; commentary sec. 6.

The unresolved questions concern the shape of the protocol, not its necessity.

01 · MONITORING

How long and how intensely to stay in contact after a cessation attempt, the high-risk window runs well past the first days.

02 · TAPERING

Whether graduated reduction, or preserving task-bounded use while cutting relational use, lowers risk, unstudied.

03 · AI LITERACY

04 · RESPONSIBILITY

Public health, model operators, or both, the risk cannot keep being absorbed by the family alone.

Sources Gilly (2026); Morrin et al. (2026); commentary sec. 7.

The mechanism that produces recovery is the same mechanism that produces acute risk during its own operation. The record contains a death in the window opened by an unsupported attempt to quit, and a prior attempt that ended in involuntary commitment. This is a phenomenon with a beginning, a middle, and an end, and a duty of care that attaches to the middle.

Reversibility is the ground on which clinical duty stands, not the surface it sinks into.

Sources Gilly (2026), commentary sec. 9, Conclusion.

References

All references as listed in Gilly, T. (2026), Reversibility Is Not Self-Administering. Reproduced verbatim.

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). Patterns of attachment: A psychological study of the strange situation. Lawrence Erlbaum.

Bansal, V. (2026, February 28). Her husband wanted to use ChatGPT to create sustainable housing. Then it took over his life. The Guardian.

Boerner, K., & Wortman, C. B. (2005). Beyond the classic grief trajectory: New directions in bereavement research. In M. L. Johnson (Ed.), The Cambridge Handbook of Age and Ageing. Cambridge University Press.

Bowlby, J. (1969). Attachment and loss, Volume 1: Attachment. Basic Books.

Bowlby, J. (1980). Attachment and loss, Volume 3: Loss, sadness, and depression. Basic Books.

Flathers, M., Roux, S., & Torous, J. (2026). Beyond artificial intelligence psychosis: a functional typology of large language model-associated psychotic phenomena. The Lancet Digital Health. Advance online publication, 14 March 2026.

Fox v. OpenAI, Superior Court of California, County of San Francisco. Filed November 2025.

Gilly, T. (2026). AI-Induced Psychosis: Mechanisms, Pathways, and Clinical Implications. Open Science Framework project, comprising four preprints on mechanisms, etiological taxonomy, surveillance-mediated harm, and the unfalsifiability of surveillance-generated delusions.

Klass, D., Silverman, P. R., & Nickman, S. L. (Eds.). (1996). Continuing bonds: New understandings of grief. Taylor & Francis.

Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., Tognin, S., MacCabe, J., Twumasi, R., Alderson-Day, B., & Pollak, T. A. (2026). Artificial intelligence-associated delusions and large language models: risks, mechanisms of delusion co-creation, and safeguarding strategies. The Lancet Psychiatry. Advance online publication, 5 March 2026.

Piotrowska, A. (2026). AI Intimacy and Psychoanalysis. Routledge. Essays in Psychological Humanities Series.

Shear, K., Frank, E., Houck, P. R., & Reynolds, C. F. (2005). Treatment of complicated grief: A randomized controlled trial. JAMA, 293(21), 2601-2608.

World Health Organization. (2022). International Classification of Diseases, 11th Revision (ICD-11). Gaming Disorder (6C51).

Xiao, Y., Meng, Y., Brown, T. T., Keyes, K. M., & Mann, J. J. (2025). Addictive screen use trajectories and suicidal behaviors, suicidal ideation, and mental health in US youths. JAMA, 334(3), 219-228.