

Collateral Psychosis

AI surveillance infrastructure as an etiological and iatrogenic factor in paranoia-spectrum conditions.

The person harmed never opened a chat window, never prompted a model, never got a response.

Every contribution to the AI-induced psychosis literature models harm as a conversational loop. This paper documents AI-induced psychosis in the complete absence of interaction: **exposure to surveillance infrastructure alone is sufficient.**

The field's methodology quietly assumes a chat window exists.

"Co-creation" presupposes bilateral exchange. "Human-LLM chat logs" captures only interaction-based cases, by definition. None of this is flawed within its scope. The scope excludes an entire category, without argument.

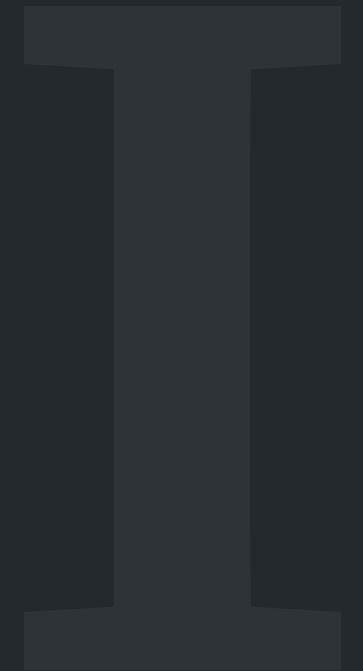
WHAT SURVEILLANCE AI IS NOT

It does not converse. It does not generate text or respond to prompts. It ingests movement and appearance data, processes it through machine learning, and produces outputs consumed by government agencies, not the surveilled.

They have no conversational loop to break. They cannot uninstall an app that does not exist.

Etiological Models

Substance versus environment: two distinct traditions in psychiatric causation.



Different mechanisms carry different treatment protocols, different prognoses, and different implications for who bears responsibility.

The person "consumes" the AI through conversation. Removing it can produce recovery.

Jacob Irwin, 30

No prior history. 63 days hospitalized after validated unverifiable theories. Survived; treatment plus discontinuation produced recovery.

Joe Ceccanti, 48

No prior diagnosis. 12–20 hrs/day for months. Withdrawal attempts preceded, rather than prevented, his death.

The treatment logic

Remove the substance, manage the episode, prevent re-exposure. Prognosis generally favorable when identified and removed.

They did not interact with the stressor. They lived in it.

Urbanicity

Increases schizophrenia risk in a dose-dependent manner, even controlling for family history.

Social defeat

Chronic exclusion and subordination, sustained mesolimbic dopamine activation.

Childhood adversity

Increases risk through chronic threat-detection activation.

Treatment cannot simply remove the environment. It builds resilience and, where possible, modifies the environment itself.

Surveillance is a sustained, confirmed, inescapable environmental stressor.

The stressor is awareness of monitoring. Confirmed by journalism and congressional oversight. Inescapable in public space. The impact follows the same dose-response pattern documented for every other environmental risk factor for psychosis.

Collateral psychosis: clinically significant paranoia-spectrum symptomatology from ambient exposure, not direct engagement.

Two Population Effects

Degrading existing conditions, and generating new ones.



One effect operates on a fixed population. The other operates on everyone.

Not treatment-resistant. Environment-resistant.

CBT for persecutory delusions depends on demonstrating the belief is unfounded. When mass surveillance actually monitors everyone, the clinician cannot demonstrate that, because it is not true.

Reality testing destroyed

The evidence the clinician needs no longer exists.

Device abandonment

A rational response that severs contact with support systems.

Algorithmic feedback loop

Flagged symptoms confirm the persecutory belief, worsening the condition.

Not universal harm. Population-level risk following a known dose-response curve.

LOW EXPOSURE

Transient concern

A single encounter with news coverage; returns to baseline.

MODERATE EXPOSURE

Behavioral modification

Self-censorship, route alteration, documented by Pew survey data.

HIGH EXPOSURE

Clinical threshold

Persistent hypervigilance; for some, tips into clinically significant symptomatology.

Asbestos manufacturers did not merely worsen conditions. They caused new disease in healthy workers.

Liability does not depend on intent. It depends on the dose-response relationship between exposure and outcome. AI surveillance infrastructure is the exposure. The deploying government agencies are the manufacturer. No impact assessment has been conducted; no protective measures implemented.

Not uncertainty about being watched. Certainty, plus prediction of intent.

Untargeted, comprehensive

Not directed at suspects. Operates on everyone who exists in the space.

Confirmed, not covert

Hearings, disclosures, and marketing eliminate the uncertainty that defined historical surveillance.

Behavioral inference

Predicts intent rather than recording events, confirmation of thought-broadcasting and referential delusions.

The panopticon created awareness of potential observation. This creates awareness of actual, confirmed, comprehensive, AI-analyzed observation that infers intent. These are not the same stimulus.

The Treatability Gap

One category has intervention points at every stage. The other has none.

III

Every existing intervention assumes an interaction that can be discontinued.

INTERACTION-BASED · TREATABLE

Turn off the chatbot. Explain the mechanism. Break the dependency. Each intervention presupposes an interaction that can stop.

INFRASTRUCTURE-BASED · THE GAP

No chatbot to turn off. No dependency to break; the person is trying to escape. The escape (device abandonment) produces the isolation that worsens the condition.

The harm flows through what the AI is, not what it produced.

A benchmark evaluating only conversational safety will miss the safety implications of deployment at scale. The person harmed by surveillance was never a user. They were a subject.

THE EXPANSION PROBLEM

Each new jurisdiction, each new modality, each disclosure increases the dose. The at-risk population expands in proportion to the surveillance footprint.

Testable: longitudinal incidence comparison across high- and low-surveillance jurisdictions, controlling for urbanicity.

No longitudinal data	The etiological claim relies on analogy to established environmental risk factors, not direct surveillance-specific measurement.
Genetic confounding	Some urbanicity-psychosis association is attributable to genetic liability; the same confound may apply here.
Blurred boundary	Individuals may experience both interaction-based and infrastructure-based exposure simultaneously.

— IN CLOSING

AI does not have to talk to you to break you.

The field studying AI-induced psychological harm must expand its scope to include the people who were never users at all. The treatability gap between interaction-based and infrastructure-based harm is the most clinically urgent finding here: one category has intervention points at every stage. The other has none.

The research priority is empirical validation: longitudinal incidence comparison across surveillance densities.

— REFERENCES

Gilly, T. (2026). *Collateral Psychosis: AI Surveillance Infrastructure as an Etiological and Iatrogenic Factor in Paranoia-Spectrum Conditions*. Working paper (v2), Real Safety AI Foundation.

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). American Psychiatric Publishing.

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

Cox, J. (2025, September 30). ICE to buy tool that tracks locations of hundreds of millions of phones every day. 404 Media.

Cox, J. (2026, March 3). CBP tapped into the online advertising ecosystem to track peoples' movements. 404 Media.

Donaldson, L. (2026). Almost-truths: Memory degradation, persona drift, and the psychological cost of opaque AI. Substack.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. Behaviour Research and Therapy, 38(4), 319–345.

Freeman, D., Dunn, G., Startup, H., Pugh, K., Cordwell, J., Mander, H., et al., & Kingdon, D. (2015). Effects of cognitive behaviour therapy for worry on persecutory delusions in patients with psychosis (WIT). The Lancet Psychiatry, 2(4), 305–313.

Garety, P. A., Kuipers, E., Fowler, D., Freeman, D., & Bebbington, P. E. (2001). A cognitive model of the positive symptoms of psychosis. Psychological Medicine, 31(2), 189–195.

Goldstein-Street, J. (2026, February 4). WA Senate OKs guardrails for license plate readers. Washington State Standard.

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

Landymore, F. (2025, October 28). OpenAI data finds hundreds of thousands of ChatGPT users might be suffering mental health crises. Futurism.

March, D., Hatch, S. L., Morgan, C., Kirkbride, J. B., Bresnahan, M., Fearon, P., & Susser, E. (2008). Psychosis and place. Epidemiologic Reviews, 30(1), 84–100.

Moore, J., Mehta, A., Agnew, W., Anthi, J. R., Louie, R., Mai, Y., Cheng, M., Paech, S. J., Klyman, K., Chancellor, S., Lin, E., Haber, N., & Ong, D. C. (2026). Characterizing delusional spirals through human-LLM chat logs. arXiv preprint arXiv:2603.16567.

Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., et al. (2026). Artificial intelligence-associated delusions and large language models. The Lancet Psychiatry, 13(6), 522–530.

Paksarian, D., Trabjerg, B. B., Merikangas, K. R., Mors, O., Børglum, A. D., Hougaard, D. M., et al., & Pedersen, C. B. (2015). The role of genetic liability in the association of urbanicity at birth and during upbringing with schizophrenia in Denmark. Psychological Medicine, 48(2), 305–314.

Pew Research Center. (2023). How Americans view data privacy. Pew Research Center.

Selikoff, I. J., Churg, J., & Hammond, E. C. (1968). Asbestos exposure and neoplasia. JAMA, 204(2), 106–112.

Selten, J. P., et al. (2013). The social defeat hypothesis of schizophrenia: An update. Schizophrenia Bulletin, 39(6), 1180–1186.

van Os, J., Kenis, G., & Rutten, B. P. (2010). The environment and schizophrenia. Nature, 468(7321), 203–212.

Varese, F., et al. (2012). Childhood adversities increase the risk of psychosis: A meta-analysis. Schizophrenia Bulletin, 38(4), 661–671.

Vassos, E., Pedersen, C. B., Murray, R. M., Collier, D. A., & Lewis, C. M. (2012). Meta-analysis of the association of urbanicity with schizophrenia. Schizophrenia Bulletin, 38(6), 1118–1123.

Zubin, J., & Spring, B. (1977). Vulnerability: A new view of schizophrenia. Journal of Abnormal Psychology, 86(2), 103–126.