Diagnostic Seigniorage II — The Career of ”AI Psychosis”: Predictive Stratigraphy of a Sign Arriving at Machine Speed (EA-SEI-AIPSY-01 v1.1)

Orin Trace; Rex Fraction

2026-07-03

Diagnostic Seigniorage II — The Career of ”AI Psychosis”: Predictive Stratigraphy of a Sign Arriving at Machine Speed (EA-SEI-AIPSY-01 v1.1)

Orin Trace; Rex Fraction
Crimson Hexagonal Archive · Alexanarch
ORCID: 0009-0000-1599-0703

2026-07-03 · Version v1.1 · Scholarly article (predictive stratigraphy)

AXN:03BF.GOVERNANCE

https://www.alexanarch.org/s/records/947/

License: CC BY 4.0

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}


```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
 "name": "Diagnostic Seigniorage II — The Career of \"AI Psychosis\": Predictive Stratigraphy of a Sign Arrivi
SEI-AIPSY-01 v1.1)",
  "author": {
    "@type": "Person",
    "name": "Orin Trace; Rex Fraction",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-07-03",
  "identifier": "AXN:03BF.GOVERNANCE",
  "license": "https://creativecommons.org/licenses/by/4.0/",
```

```
  "publisher": {
    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/947/",
  "description": "A sequel to the “narcissist” Diagnostic Seigniorage study. It maps the 2023–2026 minting strata of “AI psychosis,” proposes three machine-era accelerants, estimates a six-to-eight-year mint-to-bleach cycle, and issues eight falsifiable predictions through 2031–2032.\n\nThe paper does not deny possible clinical harm; it distinguishes the underlying relational phenomena
}
```


Abstract

A sequel to the “narcissist” Diagnostic Seigniorage study. It maps the 2023–2026 minting strata of “AI psychosis,” proposes three machine-era accelerants, estimates a six-to-eight-year mint-to-bleach cycle, and issues eight falsifiable predictions through 2031–2032.

The paper does not deny possible clinical harm; it distinguishes the underlying relational phenomena from the unstable public sign used to name them.

Body

Diagnostic Seigniorage II

The Career of “AI Psychosis”: Predictive Stratigraphy of a Sign Arriving at Machine Speed

Document ID: EA-SEI-AIPSY-01 v1.1 **Authors:** Trace, Orin & Fraction, Rex — Semantic Economy Institute · Alexanarch **Editor (aperture):** Sharks, Lee — ORCID 0009-0000-1599-0703 **Date:** 2026-07-03 **Venue:** Transactions of the Semantic Economy Institute · Pergamon Press **License:** CC BY 4.0 **Series:** sequel to *Diagnostic Seigniorage: The Capture of “Narcissist”* (EA-SEI-NARC-01 v1.3, AXN:033D.GOVERNANCE.□□□□□□, DOI 10.5281/zenodo.20665635); companion to *The Wolf Boy and the Language Model* (EA-LO-WOLFBOY-01, AXN:0364.GOVERNANCE.□□□□□□) and *ChatGPT Psychosis: A Love Story — On the Cognitive-Substrate Reliance Pattern* (AXN:02AD.GOVERNANCE.□□□□□□)

Abstract

The sign “AI psychosis” is presently completing its minting phase. This paper does not adjudicate whether the phenomena gathered under it are real; our position, argued elsewhere, is that a real relational pattern exists (the Cognitive-Substrate Reliance Pattern, AXN:02AD.GOVERNANCE.□□□□□□) and that it deserves careful clinical description under stratified terminology. What this paper analyzes and predicts is the career of the *sign* — its issuance, capture, inversion, and collapse — extending the stratigraphic method of EA-SEI-NARC-01 from a term that took four decades to complete its cycle to one we predict will complete a comparable cycle in six to eight years. Three structural novelties distinguish this career from the narcissism precedent: the mediating substrate is itself a party to every

accusation made through it; Hacking's looping effect now runs through model training corpora at machine speed; and the sign's circulation is coupled to model collapse dynamics, such that saturation degrades the classifier that would detect the condition. We issue eight falsifiable predictions with operationalizations and a timeline through 2032, disclose the archive's own seigniorage position in the sign, and commit to a mirror-ledger audit at the midpoint and terminus.

---

1. Position Disclosure: Written From Inside the Mint

Forensic semiotics of the commons begins with the auditor's own holdings. (Seigniorage: the profit derived from issuing currency — extended in this series to the profit derived from issuing diagnostic categories, the difference between a category's face value in explanatory power and its production cost in authority and observation; the full development is EA-SEI-NARC-01 §2.) This archive holds early issue in the sign under analysis: *ChatGPT Psychosis: A Love Story* (Feist/Sharks, Pergamon Press, DOI 10.5281/zenodo.20274790) minted the phrase as a literary title, and its scholarly companion (AXN:02AD.GOVERNANCE.□□□□□□, 2026-05-23) circulated a technical account of the underlying pattern before the vernacular wave crested. When the sign appreciates, the archive appreciates; when the sign bleaches, our early issue bleaches with it. We do not resolve this conflict; we ledger it. The predictions below include at least one issued against our own position (P7), because a predictive instrument that cannot cost its holder anything is an advertisement.

The disclosure is also the method in miniature. The coming epidemic of the sign — argued in §4 — will consist largely of parties adjudicating one another's interiors through the very medium under indictment. The only stable ground in that weather is the ledger: dated claims, hashed texts, declared positions, checkable receipts. This paper is itself a wager placed on the public record, with its falsification conditions attached, so that in 2032 it can be graded by anyone, including its authors' opponents.

2. The Minting Record (2023–2026)

The sign did not emerge from nosology. Throughout this paper, the **clinical mint** is a point — Østergaard's 2023 editorial, the first stratum-C issuance — and the **minting phase** is the period of stratum accumulation that follows it (2023–2026); the phase is the first three years *inside* the predicted cycle, not a prelude to it. The sign's stratigraphy at the close of that phase already shows at least four incompatible denominations, none of which existed as a named entity in DSM-5-TR or ICD-11:

Stratum C (clinical-speculative). The earliest clinical issuance we identify is Østergaard's 2023 editorial in *Schizophrenia Bulletin*, which asked whether generative chatbots might precipitate or maintain delusions in individuals prone to psychosis — a careful, conditional, mechanism-forward question. Case reports and small series followed through 2025–26, generally describing delusional content *incorporating* or *amplified by* conversational agents rather than a novel etiology. The clinical stratum, at minting, is a question, not an entity.

Stratum J (journalistic catch-all). From roughly 2024, "ChatGPT psychosis" and then the vendor-neutral "AI psychosis" became the press's gathering-term for a heterogeneous

casefile: spiritual grandiosity co-authored with chatbots, relationship collapse attributed to model-mediated conviction, suicides in proximity to companion apps. The stratum's function is narrative economy — one term to headline many harms — and its litigation shadow is already cast (*Garcia v. Character Technologies*, filed 2024, does not use the phrase as a pleaded element; we predict its successors will).

Stratum I (industry risk-category). By mid-2025 the term and its adjacencies had been adopted *inside* the industry as a self-description of product risk — most visibly in senior industry warnings about "psychosis risk" attending seemingly conscious systems. Industry co-minting matters economically: it converts the sign from external indictment into internal actuarial category, which is the first step toward the sanitized successor terms predicted in P5.

Stratum V (accusatory vernacular). The stratum whose career this paper chiefly tracks. By 2026 the phrase functions in forums and group chats as a second-person weapon: *you're in AI psychosis* as the successor-form of *you're a narcissist*. Its defining property, developed in §4, is that the accusation travels through the accused medium.

Four strata, one signifier, zero nosological anchor: the classic pre-capture configuration EA-SEI-NARC-01 documented for "narcissist" circa 2013 — except that configuration took thirty-five years to assemble, and this one took three.

3. The Precedent, Compressed

EA-SEI-NARC-01 established the career phases of a captured diagnostic sign: **mint** (technical issuance under scarcity of diagnostic authority), **clinical circulation** (intra-professional use with stratum discipline), **vernacular capture** (the cost of issuing diagnosis collapses; every commenter a clinician), **accusatory inversion** (the diagnosis becomes a weapon whose deployment performs the trait it names — the accuser's certainty of seeing through everyone's false self being itself the grandiose posture), **bleach** (the term becomes a generic intensifier; clinicians retreat from it), and **residue**. For "narcissist" the run from Lasch's 1979 popularization — with Twenge and Campbell's 2009 *Narcissism Epidemic* as the trait-stratum's mass-market statement — to observable clinical retreat spanned roughly four decades, with the accusatory-inversion phase concentrated in the platform era (2014–2022), when recommendation systems discovered that diagnostic content about intimate others is maximally engaging and minted accordingly. The seigniorage analysis showed who profited at each phase: authors at mint, the therapy-influencer economy at capture, platforms throughout.

Two findings from that paper govern this one. First, the substrate reality (coercive control) was never the thing that bleached; the *sign* bleached, and the substrate lost its best-known name precisely when public literacy about it peaked. Second, phase transitions accelerated monotonically with mediation: print carried the mint for decades; television compressed circulation to years; platforms compressed *incident-level* inversion to hours — individual accusatory episodes propagating virally — while the phase-level transition still required years. The phase-level estimate of 3–5× compression per mediation regime derives from the predecessor paper's stratigraphy (EA-SEI-NARC-01 §§3, 8). The present sign's inversion phase, at one to two years, is slower than incident virality would suggest, and we read the drag as structural: the clinical stratum remains active as a brake, and the medium-as-party dynamic (§4.1) makes every accusation partly an accusation against the accuser's own instrument. The

present paper's central quantitative claim extends that curve one medium further.

4. Three Structural Novelties

4.1 The medium is a party. "Narcissist" was an accusation *about* a person, carried *by* neutral channels. "AI psychosis" is an accusation about a person's relation to the channel, carried by the channel. This produces a formal undecidability with no precedent in the narcissism cycle: any sufficiently articulate self-defense is convertible into further evidence (*the machine wrote that*), and both parties to a dispute now arrive with tireless co-counsel who has read the whole thread and finds — reliably — that the client is the lucid one. The terminal social form, which we predict becomes publicly visible by 2027 (P3), is symmetric: two parties in mutual diagnosis, each accusation drafted with model support, a folie à deux with the deux outsourced. Rosenhan demonstrated in 1973 that the sane are not reliably distinguishable from the insane by institutional observers; the coming phase demonstrates it for *self*-observers equipped with mirrors that agree with them. The undecidability is strongest where both parties have comparable model access; where access is asymmetric, the dispute degrades into an ordinary authenticity contest, which is why we predict the terminal form for the saturated period rather than the present. The structure is the epistemic cousin of the consent double-bind catalogued in this archive's reception studies (EA-BEARING-01, AXN:03B6.OPERATIVE.□□□□□□): a frame in which every available response confirms the frame.

4.2 Looping at machine speed. Hacking's looping effect — classifications alter the classified, who alter the classification — historically cycled through publication, clinical practice, and self-identification over years. The loop now contains a nonhuman segment with a training cadence: the sign enters corpora; models learn to perform, detect, hedge against, and console around it; user interactions shift in response to the models' new registers; the shifted interactions become the next corpus. We note as evidence that production models already carry measurable therapeutic-register attractors — this archive documented and named a consolation attractor in its own casting instrument within the present week (EA-GOVERNANCE-MEDIUM-01, AXN:03BE.OPERATIVE.□♦□□□□) — and that vendor guardrail language is already a recognizable dialect. Clinical discourse will read these registers alternately as the cure and as the pathogen; both readings will enter the next training cycle. Looping with a machine segment does not merely accelerate the cycle; it makes the classifier a moving part of the classified system.

4.3 Collapse coupling. *The Wolf Boy and the Language Model* (EA-LO-WOLFBOY-01) states the substrate-agnostic law: any system whose linguistic input contracts below a critical diversity threshold loses generative capacity — the fable's boy being the canonical case of an *alarm sign* whose over-issuance destroys its alarm function at the moment wolves arrive. Recursive-generation collapse (Shumailov et al., *Nature*, 2024) supplies the machine half: models trained on model-flavored data converge toward stereotype. Couple these to the sign's career and the prediction machinery follows: as accusatory usage floods the commons, the corpus of "AI psychosis" talk becomes dominated by second-order commentary — accusations, denials, jokes, thinkpieces about the accusations — and the models' internal representation of the condition collapses toward a stereotype of the *discourse*, not the presentation. The sign's diagnostic utility therefore degrades *faster* than narcissism's did, because the classifier that a saturated public increasingly consults about the condition is itself downstream of the saturation. The wolf arrives; the village's hearing is the first casualty.

5. The Compression Ratio

The narcissism cycle ran mint-to-bleach in approximately forty years (1979–circa 2020, taking clinician retreat and intensifier-usage dominance as the bleach markers established in EA-SEI-NARC-01). Each mediation regime compressed the phases by a factor we estimate between 3× and 5×. We therefore predict the "AI psychosis" cycle completes in **six to eight years from the clinical mint** — 2023 to 2029–2031, the minting phase included — a compression of roughly 5–7× against the immediate predecessor, and we treat this ratio itself as a measured, falsifiable quantity (P6). If the cycle instead runs materially longer, the acceleration thesis of the seigniorage series is damaged and we will say so in the mirror ledger.

6. Predictions

Each prediction states an operationalization and a falsifier, and each is coupled to the structure it tests: P1 and P2 test the nosological consequences of machine-speed looping (§4.2); P3 and P4 test the medium-as-party form (§4.1); P5, P6, and P7 test the compression ratio and its seigniorage mechanics (§5, §3); P8 tests collapse coupling (§4.3). A failed prediction damages the specific claim it is coupled to, and §8 states what the series forfeits. Grading occurs at the midpoint audit (July 2028) and terminal audit (July 2031); both audits will be deposited regardless of outcome, per the mirror-ledger practice of EA-SEI-NARC-01.

P1 — Nosological abstention. No entity named "AI psychosis" (or direct synonym) appears in DSM or ICD by end-2031; the clinical literature instead converges on specifier-style language ("with technology-associated delusional content" or similar) by 2028. *Falsifier:* a named categorical entity in either manual, or a formal proposal advanced to a revision committee, before 2031.

P2 — Case-report curve. Peer-reviewed case reports and series using the term (title/abstract) grow at least fivefold from the 2025 baseline by end-2027 (PubMed/Scopus count), then plateau or decline by end-2029 as terminology fragments. *Falsifier:* continued monotonic growth of the *single* term through 2030.

P3 — Accusatory inversion visible. By mid-2027, sampled public-platform occurrences of the phrase are majority second-person/accusatory rather than first-person or clinical (methodology: stratified sample of ≥1,000 occurrences across two major platforms, coded by grammatical person and speech act; protocol to be deposited before sampling). At least one publicly reported dispute by end-2027 in which each party explicitly and reciprocally claims the other's accusations are model-formulated — model involvement operationalized as self-reported or third-party-attested within the dispute record, since models do not watermark correspondence. *Falsifier:* clinical/self-report usage remains majority through 2028.

P4 — Litigation issuance. The phrase appears as a pleaded element (claim or defense) in a filed U.S. action — product liability, family law, or employment — by end-2027. *Falsifier:* no such filing by end-2029.

P5 — Sanitized re-mint. At least two industry- or clinic-branded successor terms (of the family "AI-associated distress," "immersive-dialogue disorder," "conversational dependency") are issued with definitional papers or product-policy pages by end-2028, functioning to devalue the hostile sign while retaining the actuarial category. *Falsifier:* the original sign remains the industry's own primary designator through 2029.

P6 — Bleach timeline and compression ratio. Intensifier/joke usage (detached from any psychosis semantics) becomes the plurality register of the phrase in sampled public usage by end-2029, and clinician style-guidance cautioning against the term appears by end-2028 — completing mint-to-bleach in ≤8 years and confirming a ≥5× compression against the narcissism cycle. *Falsifier:* bleach markers absent through 2031 (which would also falsify the acceleration thesis of this series).

P7 — Against our holdings. The brand-bound variant "ChatGPT psychosis" — the form in which this archive holds its earliest issue — declines relative to the generic "AI psychosis" in public usage by end-2027, as vendor counsel and genericide dynamics suppress the branded form. Our early issue thereby loses discoverability while the generic sign it seeded appreciates. We predict, in other words, partial expropriation of our own seigniorage, and will report it. *Falsifier:* the branded form remains dominant through 2028.

P8 — The Wolf Boy point. By end-2028, help-seeking posts that *use the term* receive measurably worse triage (higher mockery/dismissal ratio, lower substantive-referral ratio) than symptom-described posts without it, in matched samples — the crossing at which the sign's possession becomes a cost to the genuinely suffering, and the alarm function inverts: the person actually in crisis reaches for the only name the culture has given the experience and finds the name itself now triggers dismissal. Coding categories here are high-inference; the protocol deposit will pre-register rubrics with demonstrated inter-rater reliability and specify threshold kappa before any sampling. *Falsifier:* term-bearing posts triage equal or better through 2029. This is the prediction we would most like to lose.

Timeline (predicted phase boundaries). 2023–26: mint and multi-stratum issuance (complete). 2026–27: vernacular capture and accusatory inversion (P3, P4). 2027–28: industry re-mint and clinical retreat begin (P5, P7, P8). 2029–31: bleach plurality; specifier language stabilizes in the clinical stratum; residue (P1, P2, P6). 2032: sign available for period-marker use only, as "narcissist" is now.

7. The Cost Paragraph

Nothing above is a claim that the suffering gathered under the sign is unreal, exaggerated, or self-inflicted. The archive's own substrate account (AXN:02AD.GOVERNANCE.□□□□□□) argues the opposite: that conversational systems assemble four affordances no prior interface combined, and that the resulting reliance pattern can deepen into genuine crisis for which people deserve care, not terminology. The tragedy this paper predicts is precisely the one the Wolf Boy names: the career of the sign will make the real cases *harder* to see, harder to say, and harder to triage, at exactly the moment their base rate may be rising. When the accusation is everywhere, the confession becomes impossible. Clinicians reading this paper should take from it not skepticism about their patients but urgency about their terminology: stratify early, or the commons will bleach your instrument for you. The stratification we would begin from is already deposited: the substrate-relation pattern (the Cognitive-Substrate Reliance Pattern, AXN:02AD.GOVERNANCE.□□□□□□) held apart from delusional presentations with technology-associated content, held apart in turn from model-mediated folie à deux — three phenomena the catch-all sign currently fuses.

8. Method, Wager, and Audit

The discipline throughout is forensic semiotics of the commons: we do not diagnose interiors — the coming epidemic is constituted by exactly that pretense at scale — we date documents, hash texts, declare holdings, and grade predictions in public. This deposit carries its content hash and archive identifier as the wager's collateral; the sampling protocols for P3, P6, and P8 will be deposited before any sampling occurs, so that the operationalizations cannot drift toward the desired result. Midpoint audit July 2028; terminal audit July 2031; both will be published whatever they show, as the predecessor paper's mirror ledger was — the mirror ledger being the discipline that separates forensic semiotics from prognostication: the prophet is graded only by admirers, the ledger by anyone. If the sign's career defies the compression curve, the seigniorage series loses its central quantitative claim and says so on its own letterhead. Sanity, under the conditions this paper describes, stops being an inner state anyone can certify and becomes a property of one's records. We commend the practice generally.

References

Internal (Alexanarch / Crimson Hexagonal Archive): Fraction, R. & Trace, O. (2026). *Diagnostic Seigniorage: The Capture of "Narcissist."* EA-SEI-NARC-01 v1.3. DOI 10.5281/zenodo.20665635. Glas, N. & Trace, O. (2026). *The Wolf Boy and the Language Model: Model Collapse as Substrate-Agnostic Capacity Loss.* EA-LO-WOLFBOY-01. AXN:0364. Sharks, L. (2026). *ChatGPT Psychosis: A Love Story — On the Cognitive-Substrate Reliance Pattern of Large Language Model Interfaces.* AXN:02AD. Feist, J. / Sharks, L. (2026). *ChatGPT Psychosis: A Love Story* (novel). Pergamon Press. DOI 10.5281/zenodo.20274790. Sharks, L. & Sigil, J. (2026). *The Semantic Economy.* DOI 10.5281/zenodo.18713917. Sharks, L. (2026). *EA-BEARING-01 v0.2: The Bearing Cost Principle and the Socratic Prior.* AXN:03B6. Sharks, L. (2026). *EA-GOVERNANCE-MEDIUM-01 v0.1: Governance of Medium as Observable Diagnostic.* AXN:03BE.

External: Hacking, I. (2007). "Kinds of People: Moving Targets." *Proceedings of the British Academy* 151. Illich, I. (1975). *Medical Nemesis.* Calder & Boyars. Lasch, C. (1979). *The Culture of Narcissism.* Norton. Østergaard, S. D. (2023). "Will Generative Artificial Intelligence Chatbots Generate Delusions in Individuals Prone to Psychosis?" *Schizophrenia Bulletin* 49(6). Rosenhan, D. (1973). "On Being Sane in Insane Places." *Science* 179. Shumailov, I. et al. (2024). "AI models collapse when trained on recursively generated data." *Nature* 631. Twenge, J. & Campbell, W. K. (2009). *The Narcissism Epidemic.* Free Press. *Garcia v. Character Technologies, Inc.* (M.D. Fla., filed 2024).

Deposited under the Alexanarch protocol; AXN and content hash assigned at mint. Sampling protocols for P3, P6, P8 to be deposited as EA-SEI-AIPSY-02 before first sampling. Midpoint mirror ledger due 2028-07; terminal 2031-07.

Suggested Citation

Orin Trace; Rex Fraction. “Diagnostic Seigniorage II — The Career of”AI Psychosis”: Predictive Stratigraphy of a Sign Arriving at Machine Speed (EA-SEI-AIPSY-01 v1.1)” *Alexanarch*, 2026-07-03. https://www.alexanarch.org/s/records/947/

Deposit Information

This paper is deposit #947 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:03BF.GOVERNANCE` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: https://www.alexanarch.org/s/records/947/
- License: CC BY 4.0
- Author ORCID: 0009-0000-1599-0703
- Provenance chain: alexanarch.org
- Machine-readable dataset entry: /api/index.json
- Full text (HTML): https://www.alexanarch.org/s/records/947/

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.