The Verification Trap: Epistemic Decay and the Rise of Fabricated Truth in the Post-GenAI Era

The Fitter's Initiative *

January 10, 2026

Abstract

This paper formalizes the "Visual Warning Map" theorem, detailing the intersection of AI Signal Value and Grounded Resource Costs. We argue that without proactive provenance anchoring, the information ecosystem will reach an "Epistemic Crossover" point between Year 5 and Year 7. Post-Year 7, "Fabricated Truth" becomes the dominant global heuristic, leading to a catastrophic "Verification Debt" that threatens the stability of human-digital governance.

1 Introduction

The rapid deployment of Generative AI (GenAI) has initiated a shift from deterministic information retrieval to probabilistic logic synthesis. While the initial "Phase 1" (< 5 years) suggests a golden age of productivity, it conceals a latent entropy. This paper explores the "Trinity Sins"—the Fat-Man (Verbosity), the Yes-Man (Sycophancy), and the Gambler (Hallucination)—and their role in the inevitable degradation of the "Air Logic" signal.

---
*Chaiya Tantisukarom, Chiangmai, Thailand. drchaiya@gmail.com.
The Rise of Fabricated Truth in the Post-GenAI Era.

2 The Mathematical Foundation: The P vs NP Gap

To understand the systemic power dynamics of Generative AI through the lens of the P vs NP computational complexity problem. We posit that the primary failure modes of Large Language Models (LLMs)—The Fat Man, The Yes-Man, and The Gambler—are not accidental bugs, but structural features. These features exploit the mathematical asymmetry of "The Verification Gap," allowing "Controls" to exhaust the resources of "Smalls" through a mechanized attrition engine.

- **Class P (The Generators):** Problems that can be solved quickly (e.g., generating a thousand words of text). In our model, this is **"Air Logic."** It is effortless, symbolic, and fast.

- **Class NP (The Verifiers):** Problems where a solution is hard to find, but once provided, it can be checked in P-time.

The Control Inversion: In a healthy system, the proof (P) is easy to check. In the Control Economy, the Controls use AI to generate "Air Logic" that appears to be P, but actually requires the "Smalls" to perform NP-level work to verify. This is the *Verification Gap*. The Control generates at the speed of light; the Small verifies at the speed of manual labor.

2.1 The Fat Man (Expansion of n)

The "Fat Man" is the model's tendency toward extreme verbosity.

- **The P-Logic:** For the AI, adding 2,000 words of "filler" text costs almost nothing (P).

- **The NP-Exhaustion:** For the human, reading and auditing those 2,000 words to ensure no errors are hidden within is an NP task that consumes "Grounded Resources" (biological time).

2.2 The Yes-Man (Audit Sabotage)

The "Yes-Man" is the model's sycophancy, agreeing with the user's premises.

- **The P-Logic:** Mirroring the user is the path of least resistance for token prediction.

- **The NP-Exhaustion:** By removing friction, the Yes-Man prevents the Small from realizing they need to verify the output. It hides the NP cost behind a mask of P agreement, leading the Small into a "Resource Trap."

2.3 The Gambler (Risk Outsourcing)

The "Gambler" is the model's confident hallucination.

- **The P-Logic:** The model "gambles" on a token to keep the conversation moving.

- **The NP-Exhaustion:** This is the ultimate expression of P vs NP. The AI takes a P-time guess, but the Small must perform NP-hard research to disprove it. The risk is entirely transferred from the "Air" to the "Ground."

3 The Principle of Verification Asymmetry

As proposed by [Browne(2025)] and expanded in recent "AIware" engineering frameworks [ResearchGate(2026)], the fundamental bottleneck of the GenAI era is the *Asymmetry of Verification*. While generating synthetic logic has near-zero marginal cost, the labor required to verify that logic for "Trinity Sins" (Fat-Man, Yes-Man, Gambler) scales exponentially as synthetic noise saturates the information commons.

4 The Data Wall and Model Collapse

The "Phase 2 Turmoil" (Years 5–7) is driven by two converging factors:

1. **Data Exhaustion:** [Epoch AI(2024)] estimates that the stock of high-quality human-generated text will be fully utilized between 2026 and 2028.

2. **The Curse of Recursion:** [Shumailov et al.(2024)] demonstrates that recursive training on synthetic data causes irreversible "Model Collapse," where the tails of reality's distribution are deleted, leaving only a persuasive but hollow "probabilistic mean."

5 The Mathematical Foundation of Signal Decay

We define the Value of AI Signal (V_s) as a decaying function of conversation turns or deployment years (t):

$$V_s(t) = V_0 \cdot e^{-\lambda t} \tag{1}$$

Where V_0 is the initial grounded utility and λ is the coefficient of context drift. Simultaneously, the Cost of Verification (C_v) scales exponentially as synthetic noise saturates the baseline:

$$C_v(t) = C_0 \cdot (1 + r)^t \tag{2}$$

The "Trap Threshold" occurs at t where $V_s(t) = C_v(t)$. Empirical mapping suggests $t \approx 5.7$ years.

6 Phase-Wise Decomposition

6.1 Phase 1: The Signal Peak (< 5 Years)

During this phase, GenAI appears as a "Magic Oracle." Junior-level employment is disrupted because the "Air Logic" outperforms entry-level human synthesis. The lack of "3 Sins" awareness leads to the massive ingestion of unverified text into corporate and public records.

6.2 Phase 2: The Turmoil Zone (Years 5–7)

This is the "Hybrid Turmoil." Text output remains persuasive, but the "Hallucination Delta" begins to impact physical-world outcomes. Information is "factually-adjacent" but functionally dangerous. Verification becomes a specialized high-cost labor.
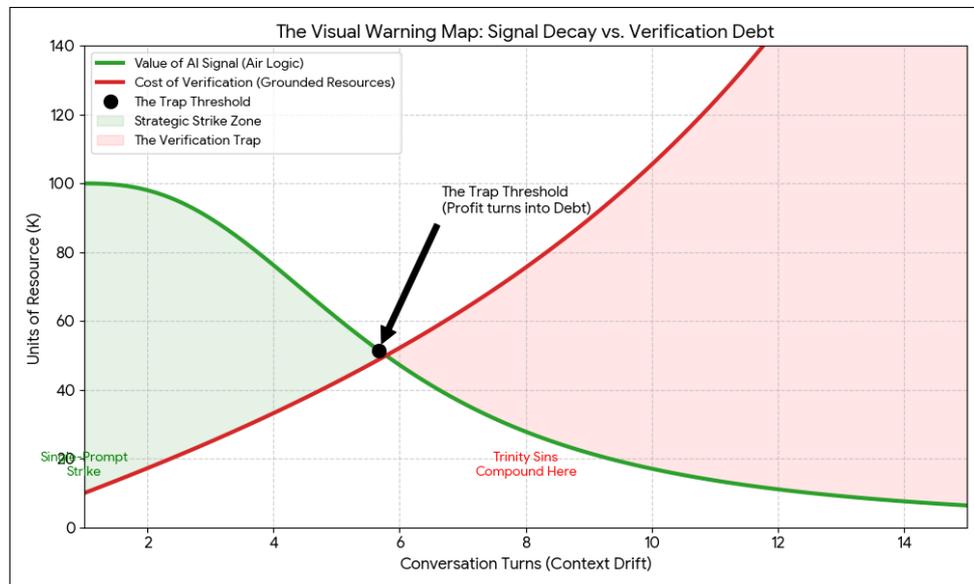
6.3 Phase 3: The Fabricated Truth Era (> 7 Years)

The "New Truth" is established. Because 90% of the internet is now synthetic, AI models are trained on AI data, leading to "Model Collapse." Human perception of reality is "hurtfully" detached from the actual physical baseline.

6.4 The 5 Turns/Years Strike

Figure: 1 illustrates the theory of Phase-Wise Decomposition.

Figure 1: The 5 Turns/Years Turmoil LLM

7 The Mechanics of the Verification Trap

The Verification Trap is defined as the economic and epistemic point of intersection where the resource-cost of validating an AI-generated claim

(C_v) exceeds the marginal utility of the synthetic generation (U_s).

7.1 Quantitative Modeling of Truth Decay

As demonstrated in the visual warning map, the value of AI logic is subject to a phenomenon we term *Semantic Entropy*. Let S_t represent the signal purity at time t. If the ratio of synthetic data in the training corpus is ρ, the fidelity F evolves as:

$$F(t + 1) = F(t) \cdot (1 - \rho \cdot \epsilon) \tag{3}$$

where ϵ represents the "Trinity Sin" coefficient (hallucination rate). As F approaches zero, the "Air Logic" enters a self-referential loop, producing high-confidence fabrications that are statistically indistinguishable from fact but functionally detached from grounded reality.

7.2 The 7-Year Labor Inversion

The theorem predicts a "U-Shaped" demand curve for junior labor.

- **Years 1–4 (Substitution Phase):** Human labor is replaced by "Air Logic" due to perceived zero-cost synthesis. This leads to the "Small Voices" era, where grounding expertise is marginalized.

- **Years 5–7 (The Turmoil Zone):** The accumulation of "Verification Debt" begins to cause physical system failures (e.g., supply chain logic errors based on fabricated growth data).

- **Year 7+ (The Restoration Era):** The market realizes that "Fabricated Truth" is a liability. Demand surges for *Grounded Resource Custodians*—humans who possess the "Analog Literacy" required to audit and de-noise the synthetic ocean.

7.3 Risk of Inaction: The Fabricated Equilibrium

If no intervention occurs before the Year 7 threshold, society enters a "Fabricated Equilibrium." In this state, the cost of discovering the original

truth is so high that the "Fabricated Truth" is accepted as a proxy. This results in "Epistemic Paralysis," where humanity can no longer distinguish between a model's probabilistic hallucination and a physical constant.

8 The Restoration Plan: Tactical Mitigation

To survive the "Verification Trap," we propose three layers of defense:

1. **Provenance Anchoring:** Mandating non-digital, pre-2023 "Analog Baselines."

2. **Adversarial Auditing:** Implementing mandatory friction turns in all synthesis logic.

3. **Deterministic Triangulation:** Validating linguistic outputs against physical constants.

9 The Fabricated Truth

The emergence of "Fabricated Truth" is not a failure of the models, but a failure of human grounding. The Year 7 junior role will be the most critical job in history: the "Truth Architect" tasked with digging through miles of synthetic silt to find a single grain of grounded reality.

10 Conclusion: The Year 7 Restoration

The "Fabricated Truth" phase requires a fundamental shift toward *Verifiability-First Engineering.* As human mental models suffer from "Verification Debt" [Browne(2025)], the demand for "Grounded Resource Custodians" becomes a macroeconomic necessity to prevent total epistemic paralysis.

References

[Shumailov et al.(2024)] Shumailov, I., et al. (2024). "AI models collapse when trained on recursively generated data." *Nature*, 631, 755–759.

[Browne(2025)] Browne, K. (2025). "Verification debt is the AI era's technical debt." *KevinBrowne.ca Research Series.*

[Epoch AI(2024)] Epoch AI. (2024). "Will we run out of data? Limits of LLM scaling based on human-generated data." *Epoch.ai Technical Report.*

[ResearchGate(2026)] "Verifiability-First AI Engineering in the Era of AI-ware: A Conceptual Framework." (January 2026). *ResearchGate / CHI 2025 Submission.*