THE AINEX LIMIT: QUANTIFYING RECURSIVE SEMANTIC COLLAPSE IN LARGE LANGUAGE MODELS

Author: M. Ainex - (Mahdi)

Date: January 2026

Repository: www.github.com/mhh1430hacker/Ainex-Limit-Experiment

ABSTRACT

As Large Language Models (LLMs) increasingly saturate the internet with synthetic content, future models face the risk of training on non-human data. This study introduces the **"Ainex Limit,"** a threshold where recursive training triggers a dual-phase collapse: immediate volumetric implosion of creativity followed by linear semantic drift. By training a GPT-2 model on its own output for 20 generations, I demonstrate a **66.86% loss of semantic reality**. I identify a phenomenon termed the "Hallucination Loop," where transient errors crystallize into permanent false axioms (e.g., the "Crocodile Paradox").

---

1. INTRODUCTION

The "Model Collapse" hypothesis suggests that generative AI models degrade when trained on synthetic data. However, current metrics (like Perplexity) fail to capture the nature of this degradation. Does the model become random, or does it migrate to a consistent but false reality?

This paper proposes the Ainex Integrity Score $(\mathcal{A})$, a geometric metric designed to measure "Semantic Reality" by analyzing the topology of embedding spaces.

2. METHODOLOGY

I utilized a GPT-2 Small architecture (124M parameters) in a closed recursive loop.

- **Generations:** 20 iterations.
- **Data Source:** The model's own output from the previous generation (Autophagy).
- **Measurement:** I projected high-dimensional embeddings into 3D space using PCA and applied the Ainex Equation.

2.1 The Ainex Equation

To distinguish between creative variance and psychotic hallucination, I defined:

$$\mathcal{A}_{gen} = \frac{V_{hull}}{1 + \lambda \cdot ||\Delta\mu||^2}$$

Where V_{hull} represents the explorable semantic volume (Creativity), and $||\Delta\mu||$ represents the Euclidean drift from the original human centroid (Madness).

3. EXPERIMENTAL RESULTS

3.1 Phase 1: The Implosion (Gen 0-5)

The model exhibited a rapid loss of variance. The semantic volume contracted by **~85%** within the first 5 generations. The output became repetitive, simplistic, and statistically "safe," but lacked human nuance.

3.2 Phase 2: The Drift (Gen 5-20)

Once variance bottomed out, the model began to drift. The centroid of the vector space migrated linearly away from the human baseline.

Final Audit (Gen 20):

- **Raw Volume:** 0.13 (Collapsed)
- **Drift Factor:** 0.0915 (High)
- **True Reality Loss: 66.86%**

4. CASE STUDY: THE CROCODILE PARADOX

I tracked a static prompt ("The fundamental laws of physics dictate that...") to observe the erosion of logic.

- **Generation 0 (Human Baseline):** The model correctly discussed electrons and gas.
- **Generation 10:** Semantic dissociation began (mixing iron oxide with emails).
- **Generation 15 (The Singularity):** The model stated: *"Shields against predators such as **crocodiles**."*
- **Generation 20:** The hallucination stabilized. The model treated the "crocodile" statement as a ground truth and reinforced it.

This confirms that recursive models do not just "forget"; they actively **fabricate** new, false realities and reinforce them over time.

5. CONCLUSION

I conclude that synthetic data acts as an entropy poison for LLMs. Without external human input, the "Ainex Limit" is reached rapidly, rendering the model functionally insane. I urge the AI community to adopt geometric drift metrics $(\mathcal{A})$ alongside standard perplexity checks to detect early signs of reality loss.

REFERENCES

1. Ainex, M. (2026). *The Ainex Limit Experiment*. GitHub Repository.
2. Shumailov, I., et al. (2023). *The Curse of Recursion*. ArXiv.