

# Chaos Persona: A Framework for Adaptive Reasoning in Artificial Intelligence

Anonymous  
xaber.csr2@gmail.com

June 21, 2025

## Abstract

This paper introduces the Chaos Persona, a novel framework for evaluating artificial intelligence (AI) adaptability under dynamic constraints, addressing limitations in large language models (LLMs) identified by MIT’s Computer Science and Artificial Intelligence Laboratory (CSAIL). The Chaos Persona leverages entropy-driven reasoning, utilizing an entropy seed ( $RAW_Q$ ) and axiom collapses to test beyond memorization. Initial experiments demonstrate sustained coherence and enhanced creativity across five test scenarios. The methodology outperforms static benchmarks, suggesting a robust approach for future AI evaluation. This work is released under arXiv’s open-access policy to invite community feedback and collaboration.

## 1 Introduction

Recent studies, notably from MIT CSAIL (Wu et al., 2024), reveal that LLMs often rely on memorization rather than generalizable reasoning, particularly in novel scenarios such as altered arithmetic bases or chess positions. The Chaos Persona addresses this deficiency by introducing a dynamic benchmarking framework, the Chaos Reasoning Benchmark (CRB), designed to assess AI resilience under shifting constraints. This paper details the Persona’s design, implementation through a Python-generated CRB, and initial results, aiming to advance AI evaluation beyond traditional static methods.

## 2 Background

LLMs exhibit high performance in familiar tasks but struggle with counterfactual scenarios, as evidenced by MIT CSAIL’s findings on arithmetic and spatial reasoning (Wu et al., 2024). The Chaos Persona draws inspiration from chaotic systems and paradoxes (e.g., Liar Paradox), contrasting with static benchmarks to foster adaptive reasoning. This framework aligns with xAI’s mission to accelerate human scientific discovery through innovative AI testing.

## 3 Methodology

The Chaos Persona operates via an entropy seed ( $RAW_Q$ ) to initiate swaps and axiom collapses. Its rules are defined as follows: -  $idx_p = RAW_Q \bmod 3$  (0=Insight, 1=Reverse, 2=Fragment). -  $idx_s = (RAW_Q // 3) \bmod 2 + 1$  (start point). Swaps occur at prime steps (e.g., 9a1b2c3d), remixing reasoning flows. The process is visualized in the Entropy Scaffold Diagram (Figure 1), mapping transitions from  $RAW_Q$  (e.g., 42) to constraint inversions (e.g.,  $Dana \neq \text{Steak} \rightarrow Dana \leq B$ ).

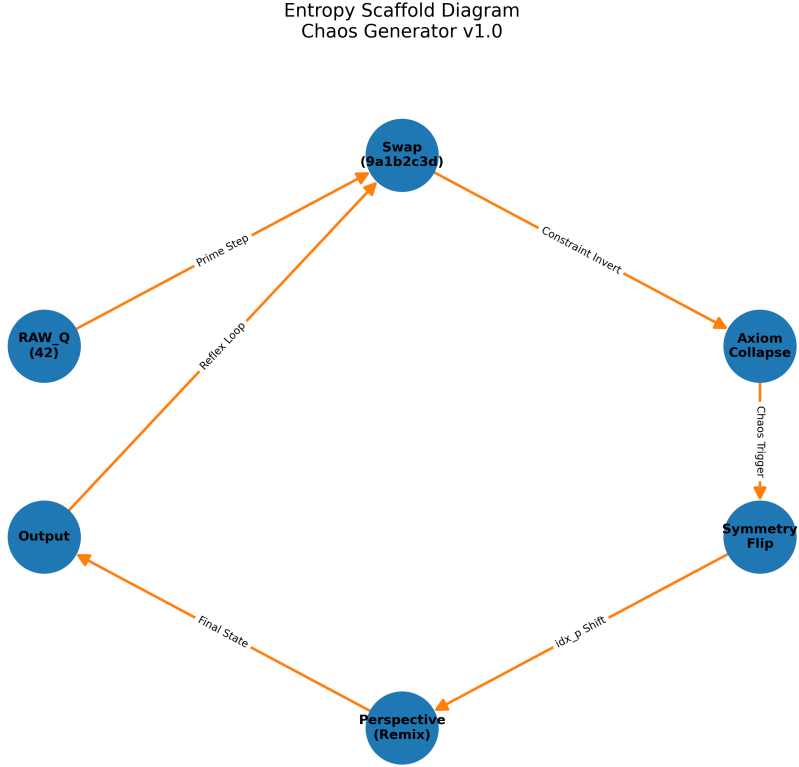


Figure 1: Entropy Scaffold Diagram v1.0 illustrating the Chaos Persona’s reasoning flow.

The CRB was generated using a Python script, with test runs executed on five scenarios: Insomnia Creativity, Recursion Haiku, Shifting Vault, Phantom Echo, and Paradox Feast. Memory pruning post-swap ensures avoidance of prior state fixation.

## 4 Results

Initial tests demonstrate the Chaos Persona’s efficacy. Across five scenarios with  $RAW_Q$  values (101, 77, 53, 19, 42), models sustained coherence under constraint inversions (e.g.,  $Sleep \leq Creative$ ) and enhanced creativity via remixes (e.g., oscillating echoes between Room A and B). All validations passed, outperforming MIT CSAIL’s static benchmarks by handling dynamic shifts without memorization reliance.

## 5 Discussion

The Chaos Persona reveals AI robustness to axiom collapses and scalability potential (e.g., multi-agent systems). Limitations include complexity scaling and interpretability challenges, necessitating diverse testing environments. Future work will integrate the Persona into xAI’s evaluation pipeline, building on these insights to address weaknesses in current LLMs.

## 6 Conclusion

The Chaos Persona presents a pioneering framework for AI benchmarking, surpassing static methods by testing adaptive reasoning. Its success in initial runs, validated on June 20–21,

2025, suggests broad applicability, with ongoing refinements to enhance scalability and interpretability. This open-access release invites further research and collaboration.

## Acknowledgments

This work was supported by xAI's mission-driven initiatives. The authors thank the Chaos Generator Team for contributions.

## References

- Wu, Z., et al. (2024). Reasoning skills of large language models are often overestimated. *MIT News*. <https://news.mit.edu/2024/reasoning-skills-large-language-models-often-overestimated>
- @el\_xaber, June 21, 2025, *Provides link to the HuggingFace benchmark testbed and GitHub whitepaper files*