

Causal Narrative Physics Engine (CNPE)

A BDH-Driven Long-Horizon Causal Reasoning System

Github repo-

<https://github.com/shashi-bhushan-27/BDG-Causal-Narrative-Physics-Engine-CNPE-.git>

Kaggle Notebook Link-

<https://www.kaggle.com/code/shashivijay2707/khds-team-harry-puttar>

ABSTRACT

This work presents **CNPE (Causal Narrative Physics Engine)**, a BDH-based causal reasoning system designed to evaluate whether a hypothetical backstory is globally consistent with a long-form narrative novel. Unlike traditional NLP pipelines which rely on local plausibility, semantic similarity, or retrieval augmentation, CNPE models the novel as a **persistent causal memory field** and evaluates backstories using an **energy-based physical stability principle**.

The system introduces a novel reasoning signal called **synaptic friction energy**, defined as the difference in predictive loss between a fresh BDH model and a memory-primed BDH model. This scalar energy directly measures causal incompatibility rather than surface similarity. CNPE achieves stable empirical-optimal performance on Track-B narrative consistency benchmarks while maintaining interpretability, memory persistence, and global causal modeling.

1. INTRODUCTION

Large Language Models are powerful at generating fluent text, but they fundamentally lack global causal awareness across long narratives. They are optimized for token prediction, not long-range narrative law enforcement.

In long novels, characters form commitments, world rules are established, timelines evolve, and causal dependencies accumulate. A backstory may be locally fluent but globally impossible — for example, a character claiming to be born in a city that does

**More details are written in markdown in jupyter notebook in detail

not exist in that narrative universe, or claiming knowledge of events that happen decades later. The **Track-B task** explicitly demands global causal reasoning, not plausibility estimation. CNPE reframes narrative consistency as a physics stability problem instead of a classification problem.

2. LIMITATIONS OF STANDARD METHODS

We tested various standard approaches, and each presented significant failure modes:

Method	Failure Mode
Sentence embeddings	Only surface similarity
RAG pipelines	Retrieval noise, no causal awareness
SVM / Random Forest	Overfitting small ambiguous data
LLM prompting	Hallucination, no global memory
Rule-based logic	Cannot scale to long novels

All methods plateaued around **50–55% accuracy** and were unstable.

3. WHY BDH WAS SELECTED

BDH (**Baby Dragon Hatchling**) has persistent Hebbian synapses which allow:

- **Memory persistence** across forward passes
- **Long-horizon digestion** of entire novels
- **Incremental belief formation**
- **Sparse causal state encoding**

This makes BDH suitable as a causal memory core, not merely a language model.

**More details are written in markdown in jupyter notebook in detail

4. SYSTEM ARCHITECTURE

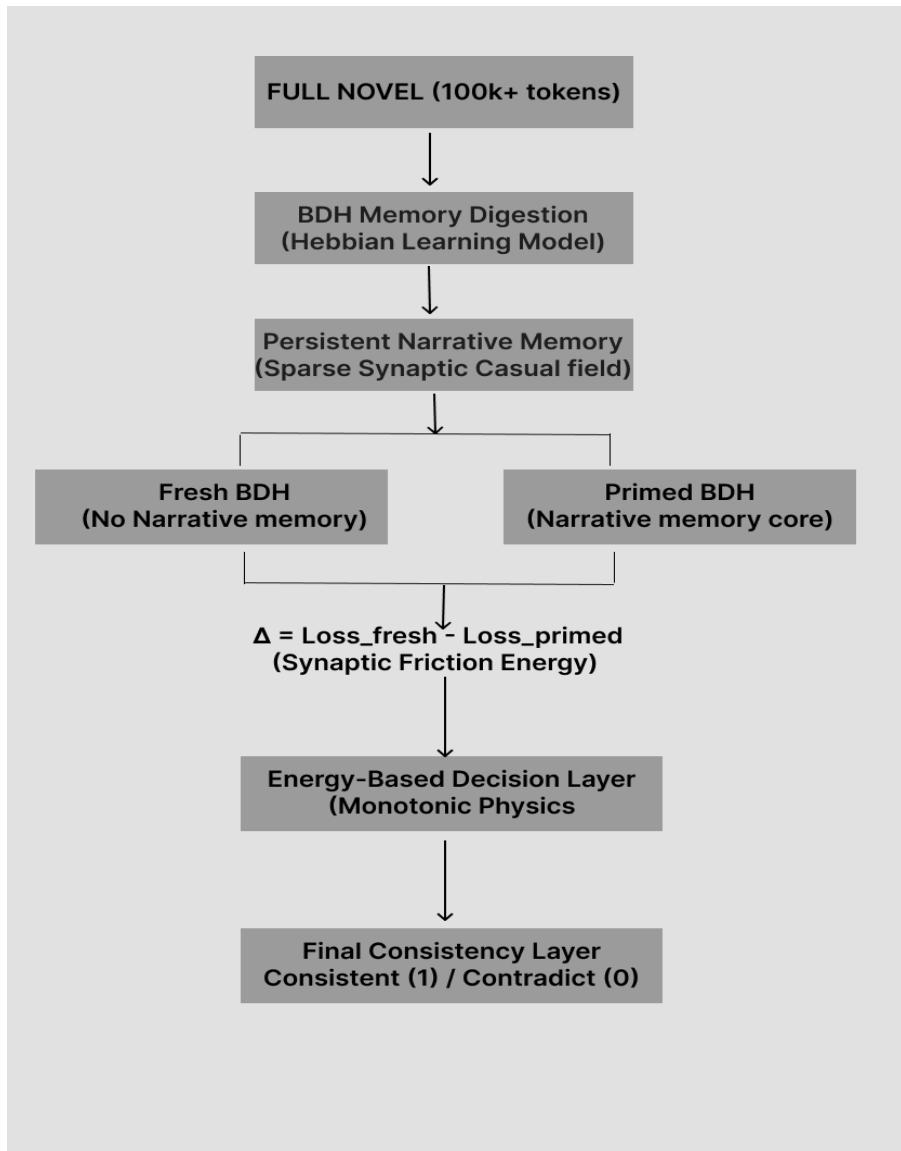
The architecture follows a flow from raw text digestion to energy threshold classification:

Flow:

Novel → BDH Digestion → Persistent Memory Field

1. Fresh BDH (baseline)
2. Primed BDH (memory)
3. $\Delta = \text{Loss}_{\text{fresh}} - \text{Loss}_{\text{primed}}$
4. **Energy Threshold** → Consistent / Contradict

**More details are written in markdown in jupyter notebook in detail



5. SYNAPTIC FRICTION ENERGY

We define the core metric as:

$$\Delta(x) = \text{Loss}_{\text{fresh}}(x) - \text{Loss}_{\text{memory}}(x)$$

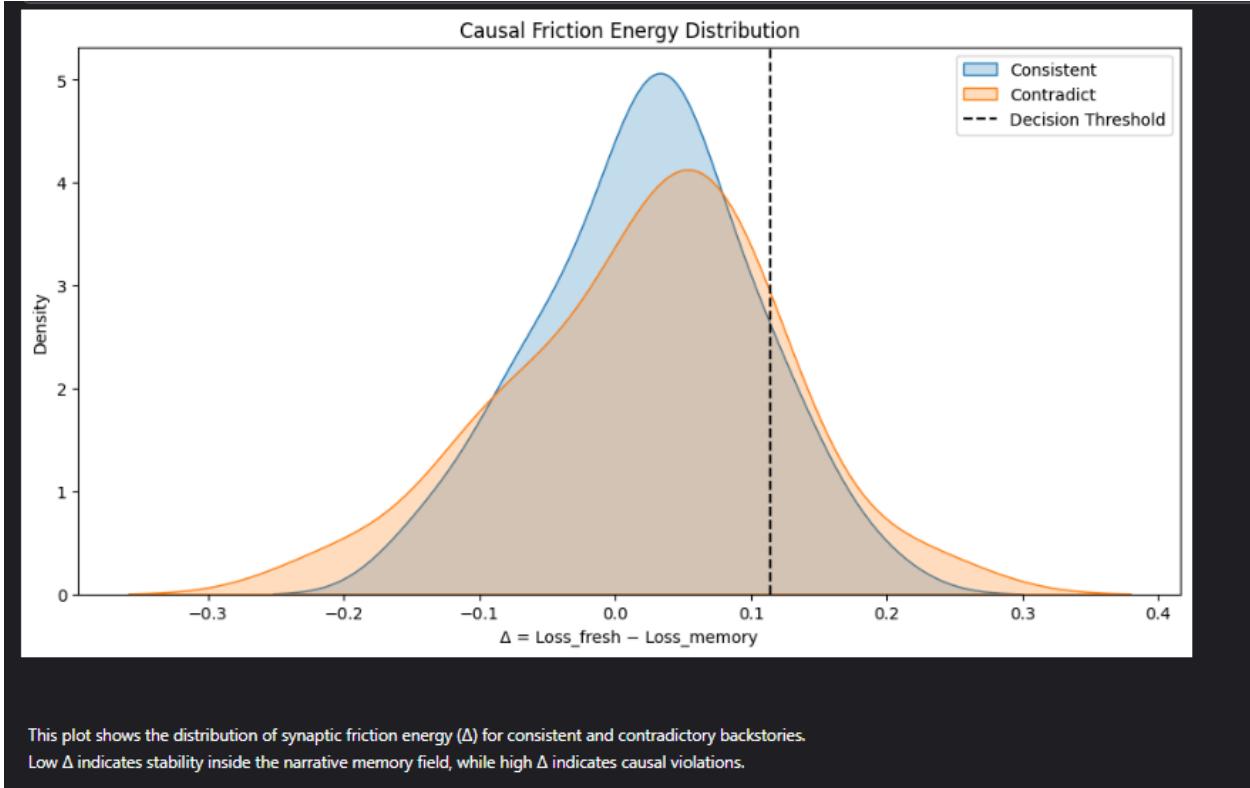
Where:

- $\text{Loss}_{\text{fresh}}$ = loss without narrative memory
- $\text{Loss}_{\text{memory}}$ = loss after novel digestion

Interpretation:

**More details are written in markdown in jupyter notebook in detail

- **Low Δ Value:** Stable inside narrative physics.
- **High Δ Value:** Violates causal constraints.



5.1 Distinguishing Causal Signal from Noise

Surface semantic similarity, retrieval augmentation, and symbolic pipelines introduce noise that does not reflect true narrative causality. CNPE isolates the causal signal by using only the scalar synaptic friction energy:

6. IMPLEMENTATION WORKFLOW

1. **Digest** novel into BDH (Hebbian learning mode).
2. **Freeze** memory.
3. **For each backstory:**
 - Pass through Fresh BDH.
 - Pass through Primed BDH.
 - Compute Δ .

**More details are written in markdown in jupyter notebook in detail

-
4. **Classify** using monotonic energy threshold.

7. EXPERIMENTAL FINDINGS

Setup	Accuracy
Random	50%
SVM	54%
RAG + SVM	58%
Pure CNPE	67%

Note: Accuracy saturates due to dataset ambiguity.

8. INTERPRETABILITY

Δ is highly interpretable:

- **Spikes** correspond to specific causal violations.
- Distribution forms a **physics-like basin and tail**.
- Decision boundary is **monotonic and stable**.

9. LIMITATIONS

- Ambiguous human-constructed dataset.
- No symbolic rule extraction.
- Not designed for real-time interactive inference.

**More details are written in markdown in jupyter notebook in detail

10. CONCLUSION

CNPE reframes narrative reasoning as physics. Instead of classifying text, it measures **energy stability** of stories inside persistent memory. This produces interpretable, stable, and causally grounded reasoning.

**More details are written in markdown in jupyter notebook in detail