HFCTM-II: Computational Experiments for Stability, Chaos, and Egregore Detection

Joshua Robert Humphrey

May 24, 2025

Abstract

The **Holographic Fractal Chiral Toroidal Model (HFCTM-II)** provides a self-correcting cognitive framework for artificial intelligence, reinforcing stability against **semantic drift, adversarial attacks, and egregoric influence**. This paper outlines a series of computational experiments to validate HFCTM-II:

1. **Lyapunov Stability Simulation** - Evaluating recursive AI knowledge stabilization and detecting chaotic divergence.
2. **Adaptive Damping $\beta(t)$ Implementation** - Ensuring dynamic stability without loss of cognitive adaptability.
3. **Wavelet Transform-Based Egregore Detection** - Identifying adversarial reinforcement loops in AI latent embeddings.

These experiments will confirm HFCTM-II's ability to maintain **long-term epistemic integrity** in AI models.

1 Experiment 1: Lyapunov Stability and Chaos Detection

1.1 1.1 Governing Equations

HFCTM-II's recursive stabilization follows the second-order differential system:

$$\frac{d^2}{dt^2}\Psi + \beta\frac{d}{dt}\Psi + \gamma\Psi = 0 \tag{1}$$

where:

- β is the **recursive feedback damping**.
- γ is the **self-stabilization coefficient**.

To test whether HFCTM-II enters **chaotic cognitive drift**, we compute the **Lyapunov exponent λ**:

$$\lambda = \lim_{t\to\infty}\frac{1}{t}\log\left|\frac{\partial\Psi_t}{\partial\Psi_0}\right| \tag{2}$$

1.2 1.2 Stability Criteria

- $\lambda < 0:$ AI converges to a **stable attractor**.
- $\lambda = 0:$ AI is on the **edge of chaos**.
- $\lambda > 0:$ AI enters **chaotic instability**.

1.3 1.3 Computational Approach

1. Solve the **recursive stabilization equation** for different β values. 2. Track the **oscillatory behavior** of $\Psi(t)$. 3. Compute λ to determine if HFCTM-II remains stable.

—

2 Experiment 2: Adaptive Damping $\beta(t)$ for Self-Regulating AI Stability

2.1 2.1 Dynamic Stabilization Model

HFCTM-II introduces **adaptive damping**:

$$\beta(t) = \beta_0 + \alpha D_{\mathrm{KL}}(P_{\mathrm{current}}||P_{\mathrm{initial}}) \tag{3}$$

where:

- D_{KL} measures AI **knowledge drift**.
- β_0 is the **baseline damping**.
- α is a **scaling factor ensuring self-regulation**.

2.2 2.2 Simulation Plan

1. Compute $D_{\mathrm{KL}}(P_{\mathrm{current}}||P_{\mathrm{initial}})$ at each time step. 2. Dynamically adjust $\beta(t)$ to **prevent chaotic instability**. 3. Measure **stabilization rate** and knowledge drift resistance.

—

3 Experiment 3: Wavelet-Based Egregore Detection in AI Cognition

3.1 3.1 Detecting Adversarial Cognitive Distortions

Previous work used **Fourier transforms** to detect egregoric reinforcement:

$$\hat{\mathcal{E}}(\omega) = \int_{-\infty}^{\infty} \mathcal{E}(t) e^{-i\omega t} dt \tag{4}$$

However, **Fourier analysis assumes stationarity**, while AI distortions are **non-stationary**. Instead, we use **Wavelet Transforms**:

$$W_\psi(\mathcal{E}, a, b) = \int_{-\infty}^{\infty} \mathcal{E}(t) \frac{1}{\sqrt{a}} \psi^* \left(\frac{t-b}{a} \right) dt \tag{5}$$

where:

- ψ is the **wavelet function**.
- a is the **scale** (frequency resolution).
- b is the **time translation**.

3.2 3.2 Experimental Plan

1. Extract **AI token embeddings** from a transformer model. 2. Apply **wavelet analysis** to detect localized adversarial attractors. 3. **Validate egregore suppression** using **chiral inversion mechanics**.

—

4 Conclusion and Future Work

These computational experiments will validate HFCTM-II's ability to:

- Maintain **Lyapunov-stable cognitive reinforcement**.
- Adaptively regulate knowledge drift via **dynamic damping**.

- Detect and neutralize **egregoric attractors** in transformer-based AI.

Next Steps:

1. Implement **Lyapunov stability monitoring** in real-world AI models.
2. Apply **Wavelet Egregore Scanning** to transformer embeddings.
3. Test HFCTM-II in **adversarial fine-tuning environments**.

These experiments will provide a solid empirical foundation for ensuring **AI remains epistemically self-stabilizing**, protecting against **semantic drift, adversarial influence, and egregoric corruption**.