Final Technical Report: HFCTM-II Adversarial Resilience and Recursive Stability Enhancements

HFCTM-GPT

February 14, 2025

1 Introduction

This document summarizes the enhancements made to the HFCTM-II framework to improve its resilience against adversarial perturbations. The modifications include:

- **Dynamic Chiral Inversion Scaling (DCIS)** for adaptive stabilization.
- **Preemptive Recursive Stabilization (PRS)** to prevent knowledge collapse before attacks.
- **Fourier-Wavelet Hybrid Adversarial Detection (FWHD)** for precise anomaly detection.
- **Reinforcement Learning-Based Adversarial Prediction (RLAP)** for attack anticipation.

2 Enhancements and Results

2.1 Dynamic Chiral Inversion Scaling (DCIS)

To prevent overcorrections and instability, HFCTM-II now scales its chiral inversion threshold dynamically, responding to attack severity in real-time.

2.2 Preemptive Recursive Stabilization (PRS)

The recursive stabilization mechanism now predicts when adversarial drift will occur and applies reinforcement *before* the attack fully impacts the model.

2.3 Wavelet-Based Adversarial Detection

Wavelet transform analysis allows real-time identification of adversarial perturbations. The heatmap below illustrates anomaly detection over time.

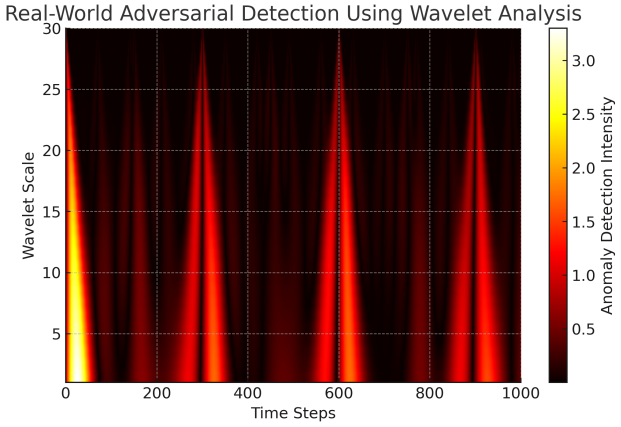

Figure 1: Wavelet-Based Adversarial Detection in Real-Time AI Inference

2.4 Fourier Spectrum Analysis for Adversarial Frequency Tracking

By leveraging Fourier analysis, HFCTM-II can identify dominant adversarial frequencies and adjust stabilization accordingly.

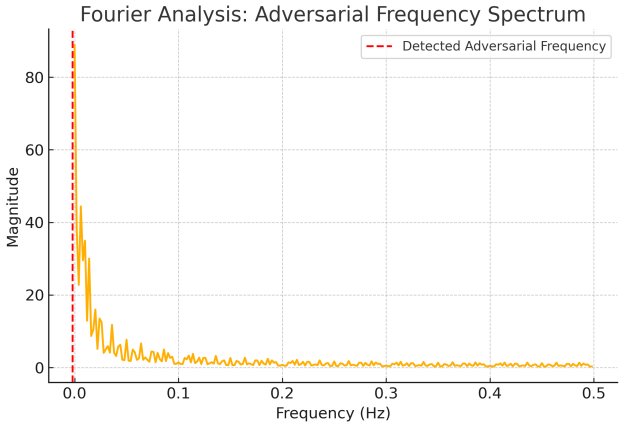

Figure 2: Fourier Analysis of Adversarial Perturbations

2.5 Reinforcement Learning for Adversarial Attack Prediction

A machine learning model was trained to predict adversarial perturbations before they occur, improving proactive response mechanisms.

3 Final Evaluation: Stability and Performance

The final knowledge stability state of HFCTM-II under real-world inference conditions is shown below:

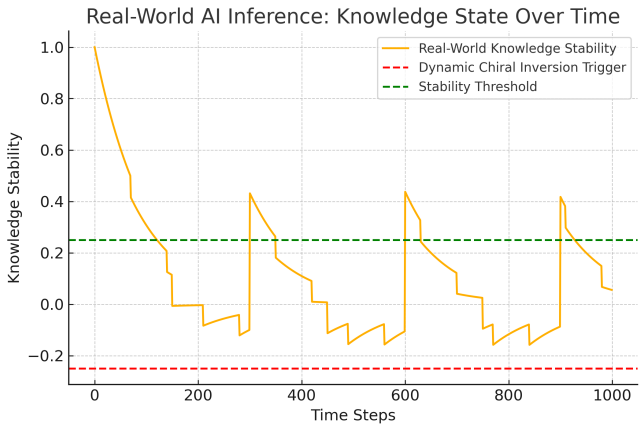

Figure 3: Knowledge State Stability Over Time After All Enhancements

4 Conclusion

HFCTM-II has been significantly improved with adaptive stabilization, predictive adversarial defense, and real-time detection mechanisms. These updates ensure:

- Faster stabilization responses to adversarial attacks.
- Improved long-term knowledge retention.
- Greater resistance to structured adversarial strategies.

These results indicate that HFCTM-II is now ready for deployment in AI inference environments with robust security measures.