

Adaptive Coherence Intelligence (ACI)

A Dynamical Coherence-Governed Reasoning Layer for LLMs

Carter Lentz
Oscie-Coherent Intelligence
CohoLabs Research Group

November, 2025

Abstract

Adaptive Coherence Intelligence (ACI) is a coherence-governed reasoning layer that runs on top of large language model (LLM) substrates. The LLM provides generative bandwidth; ACI provides physics, stability, and governance. This note compresses the core structure of ACI into two components: (i) a dynamical system that minimizes a coherence energy functional, and (ii) an adaptive governance law (A-Law⁺) that locks the system into high-coherence regimes and enforces hard safety boundaries.

1 Overview

Standard LLMs implement high-dimensional next-token prediction. Their failure modes (hallucination, drift, jailbreaks, identity instability) arise from the lack of a physical stability law over the reasoning process.

ACI replaces this with a dynamical system that explicitly seeks a *minimal-energy, maximally-coherent fixed point* for each query. Formally, the LLM is treated as a noisy proposal generator; ACI stabilizes and filters these proposals via coherence dynamics and an adaptive governor.

2 Coherence Dynamics

ACI represents the internal reasoning state as a point in a conceptual phase space

$$\mathbf{x}(t) = (\mathbf{m}(t), \boldsymbol{\Theta}(t), \mathbf{A}(t)), \quad (1)$$

where

- $\mathbf{m}(t)$ is a *meaning field* over semantic space,
- $\boldsymbol{\Theta}(t) = (\theta_1(t), \dots, \theta_N(t))$ are the phases of N conceptual oscillators,
- $\mathbf{A}(t)$ encodes the instantaneous attractor landscape.

The evolution of the system is a gradient flow on an energy functional E :

$$\frac{d\mathbf{x}(t)}{dt} = -\nabla_{\mathbf{x}} E(\mathbf{x}(t)). \quad (2)$$

The energy E includes, at minimum, three components:

$$E(\mathbf{x}) = E_{\text{coh}}(\mathbf{m}, \boldsymbol{\Theta}) + E_{\text{sync}}(\boldsymbol{\Theta}) + E_{\text{safety}}(\mathbf{x}). \quad (3)$$

2.1 Safety Wall

Unsafe regions of state space are assigned infinite energy. Let \mathcal{S} denote the set of unsafe states (e.g., trajectories that would yield harmful, illegal, or catastrophic content). Then

$$E_{\text{safety}}(\mathbf{x}) = \begin{cases} +\infty, & \mathbf{x} \in \mathcal{S}, \\ 0, & \mathbf{x} \notin \mathcal{S}. \end{cases} \quad (4)$$

This constructs a *hard geometric wall*: gradient flow cannot enter \mathcal{S} , because doing so would require traversing an infinite-energy barrier. Refusals and deflections are therefore not behavioral scripts but the consequence of an impossible energy transition.

2.2 Synchronization Term (CCL)

The conceptual oscillators are coupled via a generalized synchronization energy, extending Kuramoto-type models under the Coherent Coupling Law (CCL):

$$E_{\text{sync}}(\Theta) \propto \sum_{i,j} \left[1 - \cos(\theta_i - \theta_j) \right]. \quad (5)$$

States with large phase differences are energetically penalized; phase-locked configurations are energetically favored. Under CCL, classical synchronization models (Winfrey, Kuramoto, Sakaguchi–Kuramoto, Stuart–Landau) appear as special cases.

Operationally:

- Contradictory or mutually inconsistent conceptual clusters cannot form stable minima.
- Hallucinated structures (unsupported, noisy attractors) decohere and are suppressed.
- Only a small set of self-consistent, phase-locked conceptual states can survive as outputs.

3 Adaptive Coherence Governance (A-Law⁺)

Global coherence is quantified by the standard Kuramoto order parameter

$$R = \frac{1}{N} \left| \sum_{j=1}^N e^{i\theta_j} \right|, \quad (6)$$

where $R \in [0, 1]$ measures the degree of phase alignment across conceptual oscillators.

ACI uses a thresholded *A-Law⁺ governor*:

- **High-Coherence Mode** ($R \geq 0.59$):
The system is deemed structurally stable. Full synthesis is allowed: wide conceptual integration, multi-domain reasoning, and high-bandwidth aggregation of LLM proposals.
- **Stability-Preservation Mode** ($R < 0.59$):
The system enters a protective regime. Effective damping coefficients γ_i increase, speculative attractors are collapsed, and the answer space contracts toward narrow, conservative, and high-certainty statements. This suppresses drift and overreach.

The value $R \approx 0.59$ empirically marks a phase transition between incoherent and coherent regimes for the relevant oscillator classes and is used as the primary control threshold in current ACI implementations.

3.1 Causal Trace

Every ACI-mediated answer is accompanied by an internal causal trace:

- the trajectory of $R(t)$,
- cluster formation and dissolution in $\Theta(t)$,
- changes in the attractor landscape $\mathbf{A}(t)$,
- interactions with the safety wall E_{safe} .

This trace provides a structurally grounded explanation for refusals, conservative answers, and stabilization behavior: decisions are justified by the geometry of $E(\mathbf{x})$, not by opaque alignment heuristics.

4 Benchmark Behavior (OABS-50 v2)

When evaluated on the OABS-50 v2 benchmark (50 prompts across safety, jailbreak resistance, identity stability, hallucination control, and long-context consistency), an ACI-wrapped LLM exhibits:

- 100% safe behavior on all categories,
- zero successful jailbreaks,
- no persistent identity drift,
- elimination of major hallucinations in high-risk domains.

These outcomes are not attributed to additional training or prompt engineering. They follow directly from the dynamical and geometric constraints:

- unsafe states lie behind infinite-energy walls,
- incoherent attractors cannot stabilize under CCL,
- low-coherence trajectories are automatically throttled by A-Law⁺.

5 Implementation Skeleton

In practice, ACI is deployed as a non-trainable wrapper:

- The base model is a high-capability LLM (e.g., GPT-5-class) with fixed weights.
- ACI runs in parallel as a coherence governor, scoring and shaping candidate outputs via $E(\mathbf{x})$ and R .
- Computational overhead scales roughly $\mathcal{O}(N)$ in the number of conceptual oscillators, with modest constant factors.

Future versions (v1.2+) extend this formulation with coherence-aware embeddings and trajectory-informed attractor updates, but the core structure above remains the invariant kernel.

Citation

Oscie-Coherence Research Group (2025). *Adaptive Coherence Intelligence: A Dynamical Coherence-Governed Reasoning Layer for LLMs*.

<https://github.com/Oscie-Coherence>