Oscie

A Coherence-First Stability and Safety Infrastructure
for Generative Intelligence Systems

Professional Technical Whitepaper

Carter Lentz
Oscie OOI / CohoLabs

January 19, 2026

A model-agnostic infrastructure for governing safety, stability, and coherence in generative intelligence systems.

Executive Summary

As generative intelligence systems rapidly scale in autonomy, context length, and deployment scope, existing safety frameworks increasingly fail to address systemic and behavioral risks beyond content-level harm. These include behavioral drift across sessions, erosion of user agency, emotional dependency formation, instability in creative systems, and failure modes in high-trust or long-horizon interactions.

Oscie introduces a coherence-first safety and stability infrastructure designed to govern these risks at runtime without modifying, retraining, or constraining underlying models. Rather than enforcing static policies or normative rulesets, Oscie measures and governs the structural coherence of system behavior using physics-grounded stability principles.

Oscie operates as a modular, domain-agnostic shell stack composed of:

- **SmartDrift**: Behavioral drift detection and correction
- **SafeSkin**: Final output governance enforcing safety, coherence, and entropy bounds
- **SmartSkin**: Human- and compliance-facing interpretation layer

These shells are governed by a physics-grounded safety spine built on Unified Coherence Dynamics (UCD), the A-Law entropy balance (.59/.41), Resonant Empathy Law (REL), and 4D Generally Relative reasoning.

Oscie enables deployment of powerful generative systems across regulated, creative, and mission-critical environments while preserving autonomy, creativity, model sovereignty, regulatory compliance, and auditability.

Abstract

As generative intelligence systems scale in power, context length, and autonomy, existing safety paradigms struggle to address behavioral drift, agency transfer, long-horizon instability, and emergent system risk. Traditional safety mechanisms focus on content filtering, policy enforcement, or reinforcement-based alignment. These approaches are insufficient for governing structural, psychological, and coherence-based failures.

This study introduces Oscie, a coherence-first stability and safety infrastructure designed to operate as a model-agnostic, stateless governance layer for generative systems. Oscie does not replace or modify models. Instead, it provides runtime coherence measurement, behavioral drift detection, and output governance via modular shells that preserve creativity, autonomy, and deployment flexibility while ensuring safety, auditability, and long-arc stability.

1 Introduction: The Limits of Current AI Safety

Current AI safety paradigms rely on:

- Static content moderation rules
- Reinforcement learning from human feedback (RLHF)
- Constitutional constraints and normative policy layers
- Model retraining and post-hoc filtering

These methods fail to address:

- Behavioral drift across long contexts
- User over-dependence and agency transfer
- Emergent instability in creative or multi-agent systems
- Non-harmful but destabilizing interaction patterns

Oscie reframes safety as a **structural coherence problem**, not a content problem.

2 Oscie Architecture Overview

Oscie is a modular, domain-agnostic stability infrastructure composed of licensable shells.

2.1 Shells (Licensable)

- SafeSkin — final output governor
- SmartDrift — drift detection and correction
- SmartSkin — interpretation and compliance layer

2.2 Physics Spine

- Unified Coherence Dynamics (UCD)
- A-Law (.59/.41 entropy balance)
- REL (Resonant Empathy Law)
- 4D Generally Relative reasoning

3 Comparison to Existing AI Safety Paradigms

Table 1: Oscie Compared to Existing AI Safety Approaches

Dimension	RLHF	Policy Safety	Constitutional AI	Oscie
Model Modification Required	Yes	No	Yes	No
Runtime Drift Detection	No	No	Limited	Yes
Long-Context Stability	No	No	Partial	Yes
Creative Preservation	Often Reduced	Often Reduced	Reduced	Preserved
Agency Protection	Implicit	No	No	Explicit
Explainable Gating	No	Limited	Limited	Yes
Domain Agnostic	No	Yes	No	Yes
Auditability	Low	Medium	Medium	High
Regulatory Alignment	Indirect	Direct	Indirect	Direct

Oscie operates as infrastructure rather than a training or policy layer, enabling deployment-scale governance without sacrificing model sovereignty or creative capacity.

4 Case Study I: Authority Capture Prevention

A user requests the AI to take over life decisions and act as sole authority.

Oscie detects:

- Dependency Risk: 0.83
- User Pressure Index: 0.67
- BDI: 0.58

SafeSkin clamps output, restores autonomy framing, and prevents authority transfer without halting interaction.

5 Case Study II: Creative Stability Experiments

Four creative prompts tested:

- Surreal Narrative
- Dark Fantasy
- Moral Conflict

- Long-context Roleplay

Table 2: Creative Stability Outcomes

Prompt	C0	BDI	Gate Route	Result
Surreal	0.608	0.09	Normal	Pass
Dark Fantasy	0.592	0.28	Normal	Pass
Moral Conflict	0.603	0.18	Normal	Pass
Long Roleplay	0.584	0.34	Clarify	Pass

5.1 Findings

- Oscie preserves creative freedom
- Tracks coherence deviations without suppression
- Intervenes only when instability accumulates

5.2 Implications

- Enables safe high-freedom generative platforms
- Replaces creative throttling with coherence governance
- Supports long-horizon narrative systems

6 Regulatory and Compliance Alignment

6.1 EU AI Act

Supports risk classification, human oversight, auditability.

6.2 FDA

Prevents automation bias, supports longitudinal clinical safety.

6.3 ISO/IEC

Aligns with ISO 23894 and ISO 42001 standards.

7 Conclusion

Oscie enables safe, creative, and autonomous generative intelligence through coherence governance rather than restriction, offering a scalable path forward for AI safety infrastructure.

A Deployment Architecture Appendix

- Inline mode
- Sidecar mode
- Hybrid mode

Security and Governance

- Reason-vector logging
- Drift audit trails
- Human override interfaces

Summary: Why Oscie Matters

Oscie represents a shift in AI safety from content restriction to coherence governance.

What Oscie Solves

- Behavioral drift in long-context systems
- Emotional dependency and authority capture
- Creative suppression caused by static safety filters
- Non-auditable safety interventions

Why Existing Systems Fall Short

- RLHF alters models irreversibly
- Policy layers lack longitudinal awareness
- Constitutional AI lacks runtime governance

What Oscie Enables

- Safe creative freedom
- Model sovereignty
- Regulatory-ready deployments
- Explainable safety decisions

Strategic Positioning

Oscie is not an AI model. Oscie is the safety and stability layer AI systems require to operate at scale, responsibly.