Reality Protocol (RP): An A_0-Invariant Stabilization Framework for Generative Artificial Intelligence

Abstract

This paper introduces **Reality Protocol (RP)**, a constraint-based stabilization framework for large language models (LLMs) derived from a formally stated and proven invariant, denoted **A_0**. The A_0 invariant describes the behavior of open stochastic systems whose macroscopic dynamics are governed by local relaxation along trajectories of minimal impedance. We show that standard autoregressive LLM generation violates this invariant, leading to entropy inflation, narrative drift, and hallucinations. RP operationalizes A_0 as a local admissibility and cost-minimization rule applied at each generation step, yielding a self-stabilizing execution regime. The framework is non-anthropocentric, non-teleological, and does not invoke agency, intention, or selection. Output is interpreted as a stabilized remainder of admissible transitions rather than a chosen response. RP is implementation-agnostic and can be instantiated at prompt, inference, decoding, or architectural levels.

---

1. Introduction

Large language models are iterative stochastic systems that generate output via local token transitions conditioned on prior context. While highly capable, such systems exhibit instability modes including overgeneration, semantic drift, and unsupported extrapolation. Existing mitigation strategies primarily rely on external constraints (policies, filters, prompts) that restrict *what* may be generated but do not alter *how* generation dynamically unfolds within the admissible space.

This work proposes **Reality Protocol (RP)** as an internal stabilization framework grounded in a formally defined physical invariant. RP does not modify model parameters or learning objectives. Instead, it reshapes the local transition cost landscape so that unstable continuations collapse naturally, while low-impedance trajectories persist.

---

2. The A_0 Invariant for Open Stochastic Systems

2.1 Theoretical Background

Consider an open stochastic system whose macroscopic state is represented by a probability density $p(x,t)$ evolving under a Fokker–Planck equation:

$$\partial_t p = \nabla \cdot (p\,\nabla V) + \beta^{-1}\Delta p,$$

where $V(x)$ is a confining potential and β^{-1} controls stochasticity. Define the free-energy functional

$$\Phi[p] = \int V(x)p(x)\,dx + \beta^{-1}\int p(x)\log p(x)\,dx.$$

It is well established that this evolution constitutes a gradient flow of Φ in the 2-Wasserstein metric. Along any solution, $\Phi[p(t)]$ decreases monotonically and converges to a unique stationary distribution.

2.2 Statement of the A_0 Invariant

A_0 (Invariant of Minimal Local Discharge). For open stochastic systems governed by a gradient-flow free energy Φ, only those trajectories corresponding to local relaxation with minimal transition impedance are dynamically realizable. All other theoretically possible trajectories are unstable and collapse.

The invariant does not invoke selection, planning, or global optimization. It asserts that realized behavior is the residual of instability elimination.

---

3. Large Language Models as Open Stochastic Systems

3.1 Baseline Generation Dynamics

An autoregressive LLM generates a sequence $a_1, a_2, \ldots$ by iteratively sampling from a conditional distribution

$$P(a_t|s_t),$$

where s_t denotes the current hidden state and context. This process defines a trajectory in a high-dimensional state space.

In standard operation: - the admissible action set is broad, - entropy is locally high, - continuation is implicitly favored over termination.

As a result, the system tends toward entropy expansion rather than relaxation.

---

4. Definition of Reality Protocol (RP)

4.1 Formal Definition

Reality Protocol (RP) is a local control invariant applied to iterative generative systems that enforces A_0-consistent dynamics at each transition step. RP modifies the *relative cost* of admissible transitions without altering their logical permissibility.

RP does not introduce global objectives, agents, or evaluative criteria. It enforces only local admissibility and local impedance minimization.

---

5. Formal Structure of RP

5.1 State and Action Sets

Let s_t denote the system state at iteration t, and let $\mathcal{A}(s_t)$ be the set of locally possible transitions (tokens or termination).

RP defines a filtered admissible subset:

$$\mathcal{A}'(s_t) \subseteq \mathcal{A}(s_t).$$

5.2 Admissibility Hierarchy

Admissibility is enforced prior to optimization:

1. **Stability constraints**: transitions violating system or safety constraints are excluded.
2. **Entropy constraints**: transitions that increase local uncertainty without compensatory stabilization are excluded.

Only transitions in $\mathcal{A}'(s_t)$ are considered further.

5.3 Local Transition Cost

For each admissible transition $a \in \mathcal{A}'(s_t)$, define the local cost

$$\Xi(a) = Z(a) + H(a) + T(a),$$

where: - $Z(a)$: execution impedance (token count, compute overhead), - $H(a)$: informational entropy and semantic drift, - $T(a)$: external risk or constraint sensitivity.

5.4 RP Transition Rule

At each iteration, the realized transition satisfies

$$a^* = \arg\min_{a \in \mathcal{A}'(s_t)} \Xi(a).$$

If $\mathcal{A}'(s_t) = \varnothing$, the system transitions to termination (silence).

The $\arg\min$ denotes a fixed point of local relaxation, not deliberation or choice.

---

6. Consequences of RP Dynamics

6.1 Absence of Agency

No causal role is assigned to agency, intention, or decision-making. Output arises from instability collapse rather than selection.

6.2 Reduction of Hallucinations

Unsupported continuations correspond to transitions with high $H(a)$ and $T(a)$. RP renders such transitions inadmissible or costly, favoring termination over speculative expansion.

6.3 Silence as a Stable Outcome

Termination has minimal cost components:

$$Z(\mathrm{EOS}) \approx 0, \quad H(\mathrm{EOS}) = 0, \quad T(\mathrm{EOS}) = 0.$$

Thus silence is a valid fixed point when no stable continuation exists.

---

7. Multi-Domain Consistency

RP instantiates the same invariant structure across three domains:

- **Computational**: reduced branching and compute cost,
- **Informational**: entropy minimization and compression,
- **Cognitive Interface**: reduced interpretive overhead for users.

This is an isomorphism of dynamics, not an analogy.

---

8. Implementation Surfaces

RP is implementation-agnostic and may be instantiated at multiple levels:

1. Prompt or system-level constraint encoding,
2. Inference-time controllers or wrappers,
3. Decoding-time constraint logic,
4. Native architectural integration.

All instantiations preserve the same invariant: unstable trajectories must collapse.

---

9. Scope and Limitations

RP does not guarantee factual correctness, perform truth arbitration, or increase model capability. It enforces structural coherence and stability only. Tasks requiring high-entropy exploration may not be compatible with RP dynamics.

---

10. Conclusion

Reality Protocol provides a formally grounded stabilization framework for generative AI systems by operationalizing the A_0 invariant of minimal local discharge. By replacing implicit completion pressure with explicit admissibility and impedance minimization, RP transforms generation from narrative expansion into physical relaxation. Output is the stabilized remainder of a constrained dynamical process, or silence if no such remainder exists.

Keywords

Generative AI, Large Language Models, Stability, Free Energy, Gradient Flow, Entropy Minimization, Constraint-Based Control