Admissibility Before Optimization:
Why Long-Horizon Systems Fail Without Authorization Layers

Abstract

Modern adaptive systems are predominantly designed around optimization: actions are selected by maximizing expected utility, reward, or performance. This paradigm implicitly assumes that all actions are admissible, and that undesirable outcomes can be mitigated through better objectives, deeper planning, or improved confidence estimation.

This work argues that this assumption fails fundamentally in long-horizon systems operating under irreversible structural consequences. In such systems, the primary failure mode is not suboptimal choice, but premature or unauthorized commitment that collapses future admissible space and accumulates non-recoverable structural cost.

We introduce admissibility as a first-class architectural variable, distinct from reward, risk, or utility. Admissibility is formalized as an authorization layer that precedes optimization and governs whether actions are allowed to occur at all, based on their cumulative structural impact over internal time. This layer does not optimize, plan, or correct behavior; it regulates access to irreversible transitions.

The scope of this disclosure is intentionally narrow. Only a single architectural layer is exposed, without examples, simulations, or application-specific case studies. This is not due to lack of operational content, but reflects a deliberate boundary: the goal of this work is to establish the existence and necessity of an authorization layer as an independent structural component, not to specify a complete system.

By isolating this layer, we show that it cannot be reduced to constrained optimization, reward shaping, or state augmentation without loss of essential information. In particular, no utility function defined over outcome states can encode path-dependent irreversibility constraints that evolve over internal time. As a result, optimization alone cannot recover admissibility once violated.

The contribution of this work is therefore architectural rather than algorithmic. It identifies a missing class of governing structures that operate between reflexive adaptation and fully autonomous agency, regulating when systems are permitted to collapse uncertainty into commitment. This boundary is necessary for preserving long-horizon viability, identity continuity, and structural coherence in environments dominated by irreversibility.

1 Scope of the Disclosure

This document deliberately exposes a single architectural layer: an *authorization* interface that decides whether irreversible commitment is permitted to occur at a given internal moment of system evolution.

The disclosure is intentionally incomplete by design. It does not describe or imply the internal implementation of controllers, planners, learning loops, regime transition machinery, identity continuity operators, grounding gates, or any mechanism that would directly alter trajectories. Those components may exist in a complete system, but they are not required to understand the central point made here.

What is disclosed is strictly the structural boundary between *(i)* producing candidate actions or interpretations and *(ii)* being allowed to collapse uncertainty into an irreversible commitment. In this framing, admissibility is treated as a first-class architectural variable: it is not a penalty term, not a reward adjustment, not a confidence score, and not a constrained optimization trick. It is an explicit prohibition/permission boundary that can block an action even when that action is locally optimal under any reasonable objective.

The purpose of this disclosure is therefore twofold. First, it establishes the necessity of an authorization layer that precedes optimization in long-horizon adaptive systems operating under uncertainty and irreversibility. Second, it provides a minimal formal interface for such a layer, enabling practical integration into existing pipelines without assuming replacement of their models, solvers, or optimization methods.

To avoid ambiguity, the term *authorization* is used in its strict architectural sense: it refers to the system's ability to prevent irreversible structural operations from being executed prior to meeting admissibility conditions. This layer is not a new decision-maker. It is a boundary condition that determines when decision-making is allowed to crystallize into commitment at all.

2 Problem Reframing

Consider a system evolving over discrete internal time $\tau \in \mathbb{N}$. Let $\mathcal{S}$ denote the state space, $\mathcal{A}$ the action space, and $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ the transition operator.

Standard adaptive architectures implicitly assume that all actions are admissible and differ only in expected utility. Formally, action selection is typically defined as

$$a_\tau = \arg\max_{a \in \mathcal{A}} \mathbb{E}[U(\mathcal{T}(s_\tau, a))].$$

This formulation assumes that admissibility is either trivial or reducible to reward shaping. Equivalently, it assumes that inadmissible actions can be represented as merely suboptimal ones.

In long-horizon systems with irreversible structural consequences, this assumption fails. Some actions are not worse choices; they are structurally forbidden, regardless of their expected utility.

3 Admissibility as an Architectural Variable

We introduce an authorization function

$$G : (S, A, \tau) \to \{0, 1\}.$$

The value $G(s, a, \tau) = 1$ indicates that action a is structurally admissible at internal time τ, while $G(s, a, \tau) = 0$ indicates that the action is structurally prohibited.

The function G is not a decision rule, policy, or evaluation function. It does not select actions, rank alternatives, or compare outcomes. It solely defines whether an action is permitted to enter consideration at a given stage of system evolution.

The authorization function is not derived from reward, utility, or optimization criteria and operates independently of task objectives.

4 Admissible Action Set

The admissible action set at internal time τ is defined as

$$A_{\text{adm}}(\tau) = \{a \in A \mid G(s_\tau, a, \tau) = 1\}.$$

Any optimization, if present, is conditional on admissibility:

$$a_\tau = \arg \max_{a \in A_{\mathrm{adm}}(\tau)} \mathbb{E}[U(T(s_\tau, a))].$$

This formulation does not represent constrained optimization in the classical sense. The admissible set is not derived from the objective function, nor from feasibility constraints imposed by the optimizer.

Instead, authorization precedes optimization architecturally: optimization is permitted to operate only within a space whose boundaries are defined independently of utility, reward, or performance criteria. Unlike feasibility constraints, admissibility is not defined over state-action geometry or physical realizability, but over irreversible structural consequences accumulated across internal time

5 Internal Time

Internal time τ is not equivalent to wall-clock time. It is introduced here as an abstract accumulation coordinate indexing irreversible structural exposure.

Its evolution is represented generically as

$$\tau_{k+1} = \tau_k + \Delta\tau(s_k, a_k),$$

where $\Delta\tau$ denotes an unspecified accumulation of structural burden associated with system evolution.

No assumption is made about how $\Delta\tau$ is computed, measured, or enforced. The formulation does not imply access to internal controllers, regime mechanisms, or identity-preserving operators.

The sole role of τ in this disclosure is to provide an ordering parameter with respect to which admissibility may evolve independently of external time. Internal time is not a state variable of the transition operator $\mathcal{T}$ and does not participate in state dynamics; it indexes admissibility evolution rather than system behavior

6 Structural Cost of Irreversibility

Let $C_{\mathrm{irr}}(s, a, \tau)$ denote the structural cost associated with performing action a in state s at internal time τ. This cost is non-recoverable and satisfies

$$\frac{\partial C_{\mathrm{irr}}}{\partial \tau} \geq 0.$$

Once incurred, this cost cannot be undone by future optimization.

Admissibility enforces the condition

$$\mathcal{G}(s, a, \tau) = 0 \quad \text{if} \quad C_{\mathrm{irr}}(s, a, \tau) > \Theta(\tau),$$

where $\Theta(\tau)$ defines a structural viability boundary.

Monotonicity here does not reflect accumulation of error or energy expenditure, but the non-invertibility of structural deformation once commitment occurs

7 Non-Equivalence of Optimization and Admissibility

Theorem. There exists a class of adaptive systems for which

$$\arg\max_{a\in\mathcal{A}} U(\mathcal{T}(s,a)) \notin \mathcal{A}_{\mathrm{adm}}(\tau)$$

for all admissible utility functions U.

Proof sketch. Utility functions are defined over outcome states, whereas admissibility depends on the cumulative structural impact of actions over internal time. This impact is path-dependent and non-compressible into a Markovian or terminal representation without violating the irreversibility constraint. As a result, no utility defined over outcomes can encode admissibility boundaries. Therefore, optimization cannot recover admissibility once it has been violated.
A constructive example is any system whose admissibility depends on cumulative regime transitions or identity-preserving constraints that are not representable as terminal rewards

8 Architectural Implications

The authorization layer does not enforce goals, does not minimize risk, and does not correct behavior. It prevents irreversible transitions that collapse future admissible space.

No increase in planner depth, confidence estimation, or reward shaping can substitute for this layer because all such mechanisms operate within an already assumed admissible action space.
This layer is not a safety mechanism, alignment strategy, or risk minimization framework; it remains agnostic to objectives, values, and normative criteria

9 Conclusion

This work isolates and formalizes a single architectural claim: in long-horizon adaptive systems, optimization must be preceded by authorization.

We have shown that treating all actions as inherently admissible leads to a systematic blind spot when irreversible structural consequences are present. In such regimes, failure does not arise from incorrect evaluation of outcomes, but from the absence of an explicit mechanism that governs when commitment itself is allowed to occur.

By introducing admissibility as an authorization function evolving over internal time, we separate the question of *what should be done* from the prior question of *what is allowed to happen at all*. This separation cannot be achieved through deeper planning, improved uncertainty estimation, or richer reward formulations, because all such mechanisms presuppose an admissible action space.

The authorization layer described here does not enforce goals, optimize performance, or ensure safety in the conventional sense. Its role is more fundamental: to prevent irreversible transitions that collapse future admissible space and undermine long-horizon viability. Once such transitions occur, no amount of subsequent optimization can undo their structural impact.

The broader implication is that many observed failure modes in advanced adaptive systems are not optimization failures, but architectural omissions. Without an explicit authorization layer, systems are forced to encode irreversibility implicitly, where it becomes invisible, untestable, and unrecoverable.

Admissibility before optimization provides a minimal and necessary foundation for building adaptive systems that remain coherent, interpretable, and viable over extended horizons. It defines not a solution to a specific task, but a boundary condition for any architecture that must survive its own decisions.

License

GitHub: https://github.com/petronushowcore-mx/Admissibility-Before-Optimization
DOI: https://doi.org/10.5281/zenodo.18393487