Meta-Abstract: Architectural Limits of Long-Horizon Adaptive Systems

Maksim Barziankou, Poznań 2026
Project Petronus research

Abstract

Author's Observational View

At some point while thinking about the limits of long-horizon autonomy and the continuity of adaptive systems, I noticed something trivial yet unsettling. I was sitting still, focused on a single line of thought, maintaining a continuous internal flow. A brief vibration of a phone in my pocket pulled my attention away for an instant. Almost immediately, I returned to the same thought.

Externally, nothing had changed. I was still there, thinking about the same problem. But internally, the continuity was gone. The original stream had ended. What followed was not a continuation, but a new cognitive flow that only appeared to resume the previous one. I quickly "replayed" what had already been thought and moved forward, yet I clearly sensed that the process was no longer the same. The thought unfolded differently and led elsewhere than it would have if the interruption had never occurred.

In that moment, it became evident that a change of regime can introduce a structural discontinuity without any visible error. The same person, the same topic, the same apparent behavior — yet a different internal trajectory. This observation shaped my understanding that regime transitions carry structural consequences even when local correctness is preserved, and that continuity and identity cannot be inferred from surface behavior alone.

Meta-Abstract (Prior Art)

Long-horizon adaptive systems exhibit failure modes that cannot be explained or prevented through action-level correctness, optimization accuracy, or local safety constraints alone. Across domains, systems may remain locally correct while undergoing gradual structural degradation, loss of viability, or erosion of organizational identity, leading to abrupt or irreversible failure.

This body of work establishes prior art for a class of architectural limitations underlying such failures. Specifically, it identifies three independent but complementary dimensions that are absent or conflated in existing adaptive system architectures.

First, intervention cost is not equivalent to intervention magnitude or energy. The structural burden imposed by an intervention depends on interaction coupling and may accumulate silently even when trajectories and performance metrics remain correct.

Second, long-horizon viability is not governed solely by action selection but by the admissibility of organizational regimes. Prolonged operation within certain regimes may be structurally destructive despite locally correct behavior, and regulating viability at the level of regime permissibility is fundamentally distinct from action-level safety or control.

Third, viability and identity continuity cannot be reliably observed from within an agent's decision-making process without becoming causally entangled with behavior. Observation made causal through rewards, constraints, or optimization leads to distortion, feedback amplification, or silent degradation masked by metric preservation.

Together, these works delineate a conceptual space in which long-horizon viability, regime structure, and identity continuity are architectural properties rather than optimization targets. The contribution is not algorithmic or implementational. It establishes prior art for recognizing that durable adaptive systems require architectural separation between action, regime permissibility, and long-horizon observation, and that collapsing these dimensions into a single decision surface produces unavoidable structural failure modes.

This meta-level framing defines the boundary of the problem space addressed by the Petronus research program and clarifies why incremental improvements in control, learning, or safety mechanisms are insufficient to ensure sustained adaptive viability under real-world uncertainty.