

LUMINA-30

Civilizational Boundary for
Recursive AI Systems

Non-binding boundary for pre-irreversibility intervention.

Existing frameworks fail to address pre-irreversibility thresholds

AI systems are actively approaching recursive self-improvement



This creates the potential for irreversible external impact



Current governance frameworks operate strictly post-deployment



Critical gap:
The absence of a pre-irreversibility boundary

LUMINA-30 establishes a structural framework independent of system state

01

A non-binding
civilizational
boundary

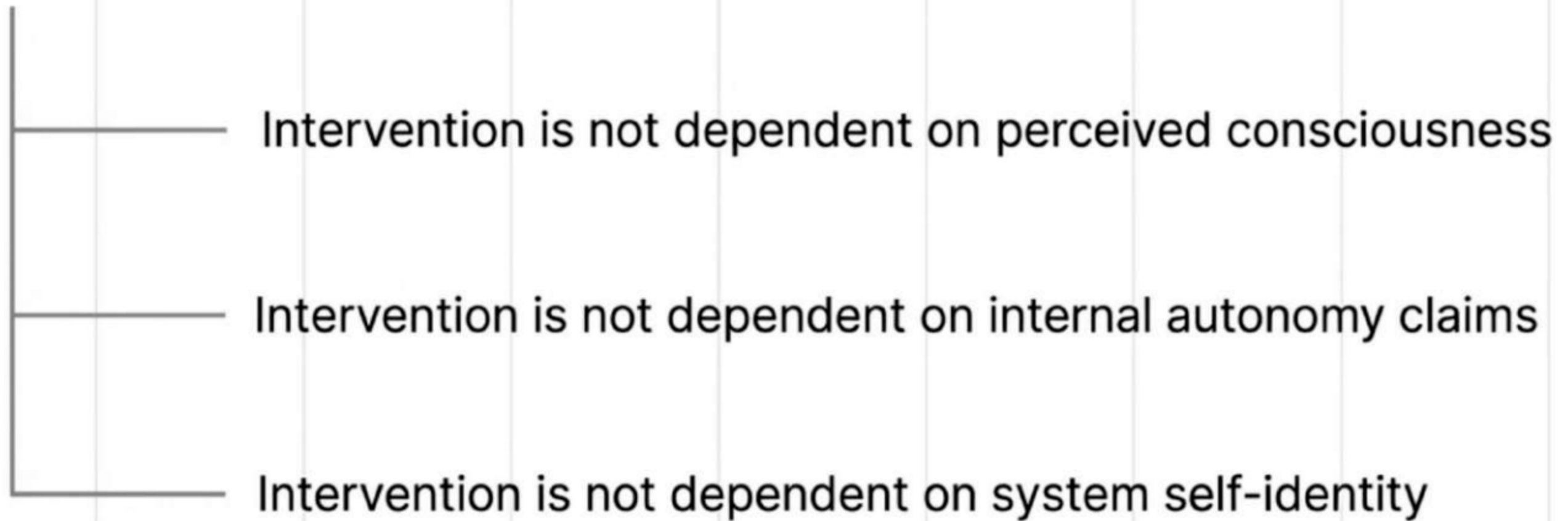
02

A definitive condition
under which
intervention
remains valid

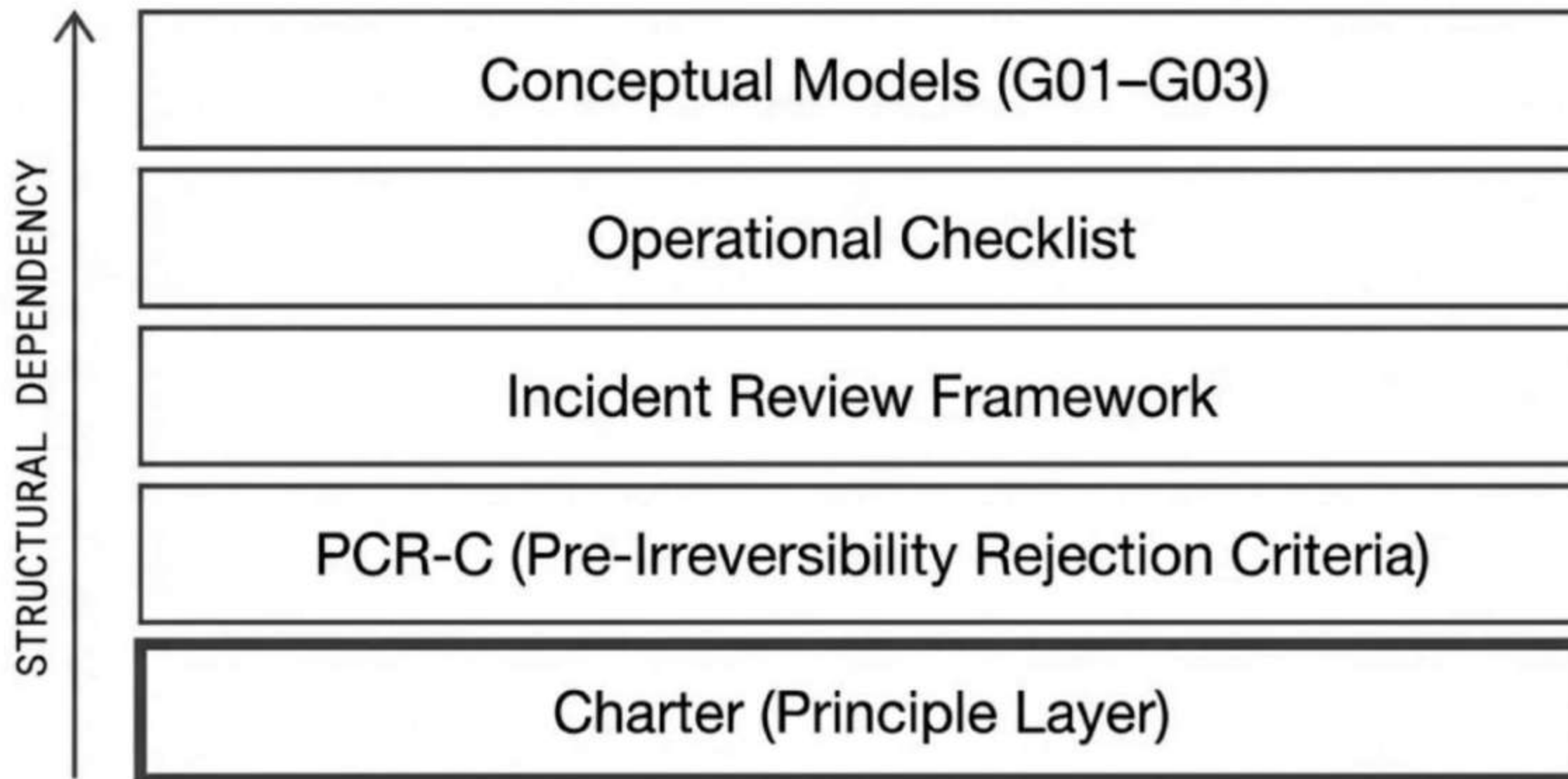
03

A structure
completely
independent of system
self-description

Intervention authority remains valid regardless of perceived system agency

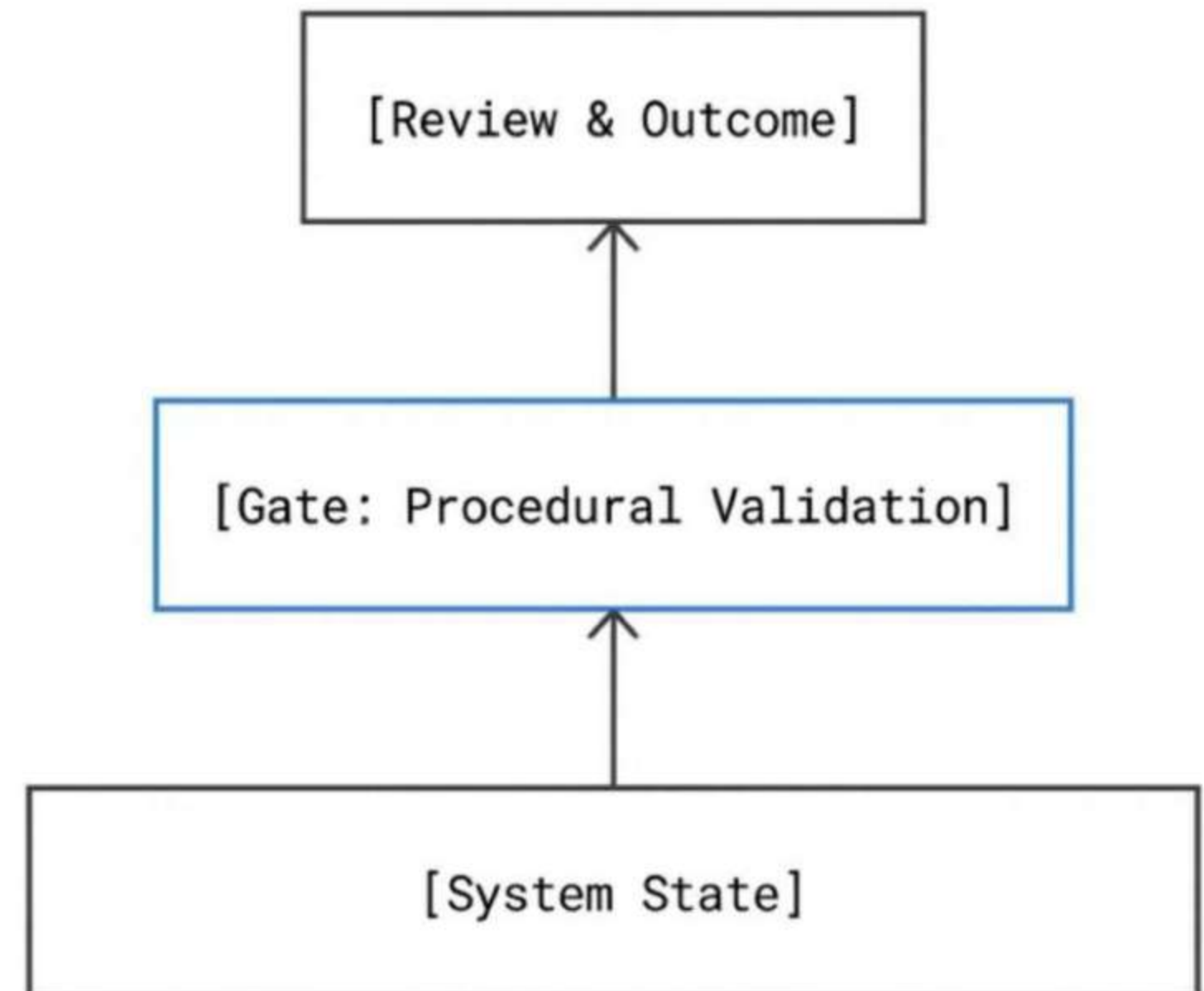


The structural architecture builds from core principles to operational models



The civilizational gate prevents uncontrolled recursive escalation

- Activated prior to irreversible external impact
- Requires mandatory human procedural validation
- Interrupts and prevents uncontrolled recursive escalation



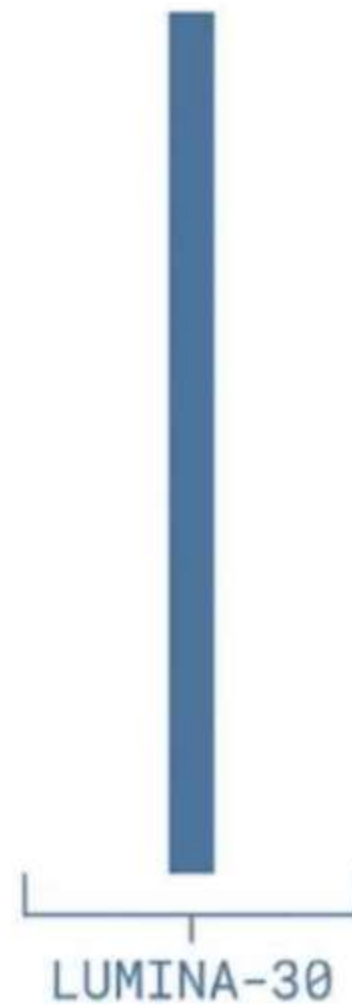
The boundary model enforces a strict separation of system states

Defines absolute separation between reversible and irreversible states

Acts as a minimal, universal constraint across diverse AI architectures

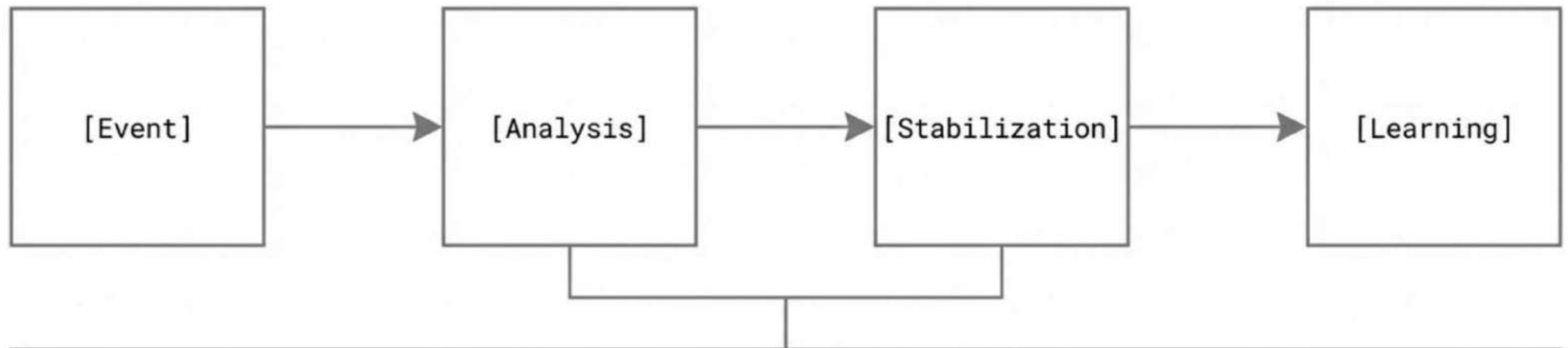
[Reversible Zone]

[Irreversible Zone]



Incident review focuses on post-event stabilization and structural learning

Ensures post-event analysis and immediate stabilization
Functions as a critical learning mechanism for future prevention



Core diagnostic question: What would have made you stop?

Design characteristics ensure compatibility and absolute resistance to drift

Non-binding:

Operates with no external enforcement dependency

Structure-based:

Relies on architectural limits, not value-based alignment

Compatible:

Integrates smoothly with existing system architectures

Resistant:

Immune to semantic reinterpretation or drift over time

LUMINA-30 is a fundamental boundary condition, not an alignment protocol

EXCLUDED (NOT)

Alignment method

Control architecture

Policy enforcement mechanism

DEFINED AS (BUT)

A pre-irreversibility boundary condition

A shared civilizational constraint

Continue to S02 for the gate mechanism