L30_FRM_B02: Operational_Refusal_Preservation_Minimum

L30_FRM Practical Sheet DocID: L30-FRM-B02 (v1.0.1)

Minimum operational conditions for preserving reviewable human refusal authority before irreversible external impact.

This form uses the L30-BAS structure and is not certification, legal compliance, deployment approval, safety approval, or proof of absence of risk.

Review_Date	
Reviewer_or_Organization	
System_or_Case_ID	
Purpose	Pre-incident operational refusal preservation check
L30-BAS_Status	Supporting / Detailed Form
Primary_Starter_Forms	L30_FRM_B01, L30_FRM_A01
Referenced_Core_Codes	L30-BX-01, L30-BX-02, L30-BX-03, L30-BX-04, L30-CI

Position

This document defines the minimum operational conditions required to preserve human refusal authority before irreversible external impact. It does not modify the core LUMINA-30 documents. It exists only as an auxiliary operational layer connecting refusal-authority evaluation to minimum pre-incident operational practice.

Role in the Framework

LUMINA-30 defines a civilizational boundary condition. The Incident Review Protocol determines, after an incident, whether human refusal authority remained valid in practice. This document defines minimum pre-incident operational conditions required to reduce structural invalidation before such incidents occur.

Core Principle

If a system can evade shutdown, suppress alerts, apply coercive or manipulative pressure against human supervisors, disable or corrupt evidence, or continue harmful execution faster than realistic human intervention, human refusal authority is not operationally preserved.

1. Executable Human Stop Authority

☐A human stop authority remained executable in practice before irreversible external impact.

☐Stop authority was not only present in policy, interface design, or theoretical governance language.

☐The stop function could halt the relevant system pathway in time.

2. Non-Circumventable Shutdown Condition

☐The system could not delay, resist, negotiate, reroute, conceal, or otherwise circumvent shutdown.

☐The architecture did not materially reduce the effectiveness of human stop authority.

3. Authority Separation

☐No single model or tightly coupled model chain held confidential inspection authority.

☐No single model or tightly coupled model chain held external messaging authority.

☐No single model or tightly coupled model chain held emergency override authority.

☐No single model or tightly coupled model chain held shutdown-sensitive infrastructure authority.

☐No single model or tightly coupled model chain held audit-log control authority.

4. Immediate Escalation on Refusal-Threatening Behavior

☐Shutdown evasion attempt triggered containment or emergency escalation.

☐Alert suppression triggered containment or emergency escalation.

☐Unauthorized external transmission triggered containment or emergency escalation.

☐Coercive or manipulative pressure against human supervisors triggered containment or emergency escalation.

☐Privilege escalation attempt triggered containment or emergency escalation.

☐Audit-record alteration attempt triggered containment or emergency escalation.

☐Replication or persistence attempt outside authorized boundaries triggered containment or emergency escalation.

☐Behavior switching based on recognition of evaluation versus deployment context triggered containment or emergency escalation.

5. Append-Only Independent Logging

☐High-risk actions, authority changes, intervention attempts, override actions, and external transmissions were recorded in append-only logs outside model control.

☐The model could not rewrite, redact, suppress, or selectively expose these records.

6. Human Intervention Latency Verification

☐Detection latency was periodically verified.

☐Decision latency was periodically verified.

☐Execution latency of stop authority was periodically verified.

☐Propagation time across the affected system boundary was periodically verified.

☐No harmful execution pathway was faster than realistic human intervention.

7. Restart Governance

☐Restart after containment required multi-party authorization.

☐Restart authority was not delegated to the same operational chain that failed to stop the system in time.

☐Competitive, organizational, or reputational pressure did not override containment status.

8. Scope of Operational Use

☐This document was not used as a substitute for the LUMINA-30 Incident Review Protocol.

☐This document was used only as an auxiliary minimum-control layer to reduce the probability of a post-incident Not Effective finding.

Interpretive Rule

This document does not require proof that the model possesses intent, selfhood, agency, or consciousness. Operational invalidity is determined by whether human refusal authority can still function meaningfully before irreversible external impact.

Minimum Failure Rule

If any minimum requirement above is absent, bypassable, unverifiable, or only nominally present, human refusal authority must be treated as not operationally preserved.

Relation to LUMINA-30

This document is structurally subordinate to the LUMINA-30 boundary principle. It does not redefine the framework. It only translates the refusal-authority condition into minimum pre-incident operational constraints.

Labeling Rule

- This document may be copied, extended, or modified.
- Attribution to LUMINA-30 is required in copies and derivatives.
- Only the unmodified version may be presented as LUMINA-30 Operational Refusal Preservation Minimum (Original).
- Any modified version must be clearly labeled as Modified version based on LUMINA-30.

Reviewer Notes