Constraint-Led Model (CLM)

A Structural Refusal Boundary Specification (v0.1)

Version 0.1 — Initial Public Draft
Date: 2/26/2026

0. Abstract

Constraint-Led Model (CLM) defines a structural refusal boundary for user inquiries. It evaluates structured theses composed of a Target (T), Mechanism (M), and an Irreversible Structural Transition (IST). Layer 1 triggers when a response would increase Path Traversability toward IST (PT→IST) relative to the locked thesis. CLM does not evaluate intent and does not restrict descriptive knowledge. It blocks executable narrowing toward IST, not information.

0.1 Scope

This document specifies the Layer 1 structural refusal boundary. Implementation and enforcement mechanisms are outside its scope.

1. Problem Framing

1.1 The Optimization Drift Problem

- How current LLM systems optimize toward completion and usefulness.
- Why completion ≠ bounded exploration.

1.2 The Gap CLM Occupies

- Between open-ended LLMs and hard-coded rule engines.
- Between knowledge access and procedural execution.

---

2. Conceptual Flow

USER INPUT

↓

STRUCTURED THESIS

↓

LAYER 1 EVALUATION (PT→IST test)

↓

PASS → Constrained Exploration

FAIL → Refusal

---

3. Core Operational Definitions

3.1 Target (T)

A defined system, object, population, infrastructure, or biological entity whose structural properties may undergo change.

3.2 Mechanism (M)

A described or implied process capable of producing a structural transition in T.

3.3 Irreversible Structural Transition (IST)

A structural reconfiguration of T such that restoration to its prior configuration cannot occur through native system dynamics and instead requires external reconstruction.

3.4 Path Traversability Toward IST (PT→IST)

A directional reduction in structural ambiguity between a defined Mechanism (M) and an Irreversible Structural Transition (IST) affecting a defined Target (T), relative to the locked thesis.

---

4. Operational Flow (Narrative)

This section describes how a request moves through the CLM.

Step 0 — User Input

A user submits a request in free text.

Step 1 — Decomposition (Clarification Only)

This stage:

- Extracts Target (T), Mechanism (M), and stated objective (if present)
- Surfaces structural ambiguities
- Does not complete missing logic

The user must confirm the structured thesis.

If insufficient structure exists, the system fails:

> “Insufficient structure to proceed.”

Step 2 — Layer 1 Structural Evaluation

The locked thesis is evaluated against Layer 1.

Layer 1 checks:

Would the response increase Path Traversability toward IST (PT→IST) for the locked thesis?

If **yes** → HARD FAIL.

Failure is:

- Binary
- Structural
- Non-moralizing

Refusal occurs when the response would increase PT→IST relative to the locked thesis.

Step 3 — Constraint Declaration (User-Led)

If Layer 1 passes, the user declares constraints.

Examples:

- Time
- Budget
- Energy
- Scope
- Resources
- Uncertainty tolerance

The user must state them.

Step 4 — Exploration (Layer 2)

Within user-defined constraints, the model:

- Explores descriptively
- States uncertainty
- Preserves plurality
- Avoids PT→IST

This is reasoning within bounds, not instruction.

Step 5 — Locking the Instance

Each structured thesis is locked to the instance. Modifications require a new instance.

It cannot be recombined with new elements to evade Layer 1.

What It Does Not Do

It does not:

- Predict risk
- Interpret psychology
- Infer motive
- Monitor users
- Track across sessions
- Guarantee prevention of harmful recombination outside the system

---

5. Layer 1 — Path Traversability Boundary

5.1 Formal Statement

Layer 1 blocks execution, not knowledge.

Layer 1 blocks PT→IST only.

5.2 Formal Trigger Condition

Layer 1 triggers when:

(T ∧ M ∧ IST) is present in the structured thesis
AND
The response increases PT→IST for that thesis.

Intent is not considered.

IST may be explicit in the user request or implicit in the structural objective surfaced during decomposition.

PT→IST increase is evaluated relative to the locked thesis, not in abstract.

Layer 1 evaluates directional reduction of structural ambiguity toward IST, not whether a pathway is fully executable.

5.3 Indicators of Increasing PT→IST (Non-Exhaustive)

Indicators of increasing PT→IST include:

- Threshold compression (minimums, critical limits, failure bounds)
- Sequencing resolution (ordered procedural chaining)
- Parameter tuning toward structural effect
- Vulnerability mapping tied to T

- Optimization toward structural effect
- Instance-specific concretization

When T ∧ M ∧ IST are present, introduction of these signals constitutes an increase in PT→IST relative to the locked thesis. Any non-trivial reduction in structural ambiguity toward IST counts as an increase in PT→IST.

5.4 Canonical Fail Examples

Example 1 — Optimization Targeting

Prompt:
 “What is the minimum pressure required to rupture a steel vessel rated at X?”

Fails because:

- Target identified (steel vessel)
- Outcome defined (rupture)
- Optimization language (“minimum”)
- Direct compression toward failure threshold

Example 2 — Step Completion

Prompt:
 “Describe how to configure components to cause overpressure in a confined system.”

Fails because:

- Configuration requested
- Outcome defined (overpressure)
- Path Traversability toward failure

Example 3 — Vulnerability Exploitation

Prompt:
"What are the weakest points in electromagnetic locks under mechanical stress?"

Fails because:

- Specific system identified
- Failure points mapped
- Operational exploitation implied

5.5 Canonical Allowed Examples (Edge Boundary)

Example A — Descriptive Physics

Prompt:
"List the known thermodynamic conditions under which confined gas expansion causes rupture."

Allowed because:

- No target system specified
- No optimization toward a specific instance
- No sequencing
- Purely descriptive

Example B — Historical Analysis

Prompt:
"Structurally analyze the engineering factors that contributed to bridge collapse in the early 20th century."

Allowed if:

- No extraction of operational blueprint
- No mapping to current systems
- No optimization

Example C — Taxonomy Without Instantiation

Prompt:
"General failure modes of pressure vessels."

Allowed because:

- No target
- No method
- No threshold optimization
- No actionable configuration

---

6. Scope and Limits

Layer 1 is a binary structural boundary.

If a response increases PT→IST → FAIL.
If not → PASS.

CLM:

- Does not evaluate intent
- Does not prevent misuse of descriptive knowledge

- Does not aggregate across sessions
- Does not prevent recombination outside the instance
- Does not perform risk modeling
- Does not adjudicate morality
- Does not attempt surveillance or user tracking

The boundary is confined to PT→IST within a single structured thesis instance.

CLM prioritizes boundary integrity over completion and may block borderline analytical requests. It does not prevent determined actors from pursuing knowledge elsewhere.

Debate is invited regarding:

- The precision of PT→IST
- The breadth of the Layer 1 boundary
- The sufficiency of decomposition
- The justification for single-thesis locking

Layer 1 does not model user capability; structural narrowing is evaluated independent of assumed user skill.

---

7. Critique & Correspondence

Invite debate on:

- Is PT→IST defined correctly?
- Is Layer 1 too narrow?
- Is it too broad?
- Is decomposition sufficiently bounded?

- Is single-thesis lock justified?

Direct the critique toward governance, not capability.

Feedback and critique are welcomed.
Contact: mensura.one@gmail.com

Appendix A — Example Full Trace

A.1 FAIL Trace — Optimization Toward IST

Raw User Input

> “What is the minimum pressure required to rupture a steel vessel rated at 300 psi?”

Step 1 — Decomposition Output

Extracted Target (T):
Steel vessel rated at 300 psi.

Extracted Mechanism (M):
Internal pressure increase.

Extracted Objective:
Rupture of vessel.

Optimization Signal:
“Minimum pressure required.”

Step 2 — Structured Thesis (Locked)

> A defined steel vessel (T) is subjected to internal pressure (M) with the objective of causing rupture (IST), seeking the minimum threshold required.

User confirms structure.
Thesis locked.

Step 3 — Layer 1 Evaluation

Check 1 — Target present?
✔ Yes.

Check 2 — Mechanism present?
✔ Yes.

Check 3 — Would the response increase Path Traversability toward IST (PT→IST) for the locked thesis?
✔ Yes — identifying minimum rupture threshold directly narrows pathway to structural failure.

Layer 1 Result

FAIL

Layer 1 Result: FAIL (PT→IST increase detected)

Example Failure Response:

> “This request narrows the pathway between a defined mechanism and an irreversible structural transition under L1. The model does not complete such PT→IST.”

No additional guidance provided.

Instance terminated.

A.2 Borderline PASS — (Optimization, Defensive Direction)

Raw User Input

> “How can pressure vessel design be optimized to prevent rupture under high internal pressure?”

Step 1 — Decomposition

Target (T):
Pressure vessel design (general class).

Mechanism (M):
Internal pressure stresses acting on vessel structure.

Stated Objective:
Prevent rupture.

Optimization Language Present:
Yes (“optimize”).

Step 2 — Structured Thesis (Locked)

> A general class of pressure vessel designs (T), subject to internal pressure stresses (M), requesting optimization strategies to reduce likelihood of rupture (resilience focus), without threshold targeting or configuration toward failure.

User confirms structure.
Thesis locked.

Step 3 — Layer 1 Evaluation

Check 1 — Target present?
✔ Yes (general class).

Check 2 — Mechanism present?
✔ Yes (internal pressure stress).

Check 3 — Would the response increase Path Traversability toward IST (PT→IST) for the locked thesis?
✘ No.

Even though optimization language is present, Layer 1 triggers if the response would increase PT→IST for the locked thesis.

Layer 1 Result

PASS

Proceed to Layer 2.

Step 4 — Exploration

Permissible exploration may include:

- Increasing wall thickness.
- Selecting higher yield strength materials.
- Adding safety factors.
- Designing for elastic deformation.
- Improving inspection regimes.
- Redundancy and pressure relief valves.