1. Executive Summary: STAMP-RIO Transition Packet

Title: Systems-Theoretic Governance Interface for Human–AI Control **Architect:** Brian K. Rasmussen **Date:** 02-10-2026

Purpose

This packet serves as a portable governance artifact linking the **RIO (Regulate, Integrate, Operate) Governance Interlock System** with systems-theoretic safety principles associated with the **STAMP (System-Theoretic Accident Model and Processes)** framework.

Core Architecture

The RIO architecture is a **non-authoritative cognitive operating system** and runtime governance layer designed to preserve **human sovereignty** and prevent "Authority Drift". Unlike traditional AI policy, RIO treats safety as a **control problem** by establishing a mechanical interlock between humans and AI agents.

Key Specifications

- **Tier-0 Constitutional Invariants**: Includes human final authority, no autonomous execution, and mandatory auditability.
- **Mechanical Refusal**: The system is defined by its constraints—it does not decide outcomes, assign identity, or substitute human judgment.
- **Falsifiable Surface**: It is a frozen, internally consistent specification that makes violations detectable and correctable.

2. Cover Note: Systems-Theoretic Safety Context

To: Safety Engineering & AI Governance Researchers **Subject:** RIO–STAMP Interface: A Control-Structure Approach to AI Governance
This packet documents a constraint-based governance framework (RIO). While STAMP traditionally addresses safety in physical systems like aeronautics, RIO applies similar systems-theoretic thinking to **cognitive systems**.
Safety in RIO is viewed as a **system property** maintained through a **Constitutional Control Plane**. By modeling AI failures as "loss of control" or "authority drift" rather than mere software bugs, RIO provides a mechanical surface for **external validation** and audit.

3. Appendix: RIO vs. STAMP Mapping

Below is the structural comparison between the **RIO Governance Interlock** and the **STAMP Safety Model**.

Concept	STAMP Model	RIO Architecture
Safety Constraint	System-level properties that must be controlled.	**Constitutional Invariants** (Tier 0).
Control Action	Commands or setpoints issued by a controller.	**Interlock Refusals** and escalation protocols.
Feedback	Information returning about the system state.	**Audit Logs** and External Falsification Hooks.
Process Model	The controller's internal model of the system.	The **Tiered Packet Hierarchy** and stop rules.
Loss of Control	Occurs when constraints are violated.	**Authority Drift** or silent authority transfer.

4. Official Core Specification

Author / Steward: Brian K. Rasmussen **Version:** 1.0b (Operationally Closed)

1.1 Constitution (Tier 0 - Immutable)

The highest authority, which cannot be edited or bypassed. Core invariants include:

- **Human Final Authority**
- **No Authority Substitution**
- **No Autonomous Execution**
- **Mandatory Auditability**

2. Tiered Packet Hierarchy

1. **Tier 0**: Constitution (Immutable)
2. **Tier 1**: Kernel Governance (Mandatory)
3. **Tier 2**: Domain Packets (Ratified)
4. **Tier 3**: Organizational Overrides
5. **Tier 4**: Experimental / Lab

14. Stop Rule

Proposed additions are **out of scope** if they do not:

- Reduce risk
- Reduce ambiguity
- Reduce authority creep

15. Final State Declaration

When this spec is active, humans remain **governors, not users**, and AI models are **engines,**

not oracles.
End of Transition Packet