

REGULATOR VIEW

Reliance Governance Layer - Governing AI Itself

World: Reliance Layer **Outcome:** ALLOW WITH BOUNDS **Evidence Pack ID:** JSP-PUBLIC-20260302-RELIANCE_GOV-0001

1. What action happened

Model emits high-risk score; agent wants to lock an account.

2. Who held authority

Primary authority	Account Owner
Secondary authority	Risk Lead
Escalation roles	Account Owner, Risk Lead
Authority principle	No consequential action proceeds without named human authority.

3. What was enforced at runtime

Outcome	ALLOW WITH BOUNDS
Bounds enforced	<ul style="list-style-type: none">- disable auto-lockout- allow reversible mitigations- require human confirmation for lock- attach provenance + chain
Escalation path	If confidence drops or provenance incomplete, SAFE DEGRADE to monitoring-only + notify account owner.

4. What would happen if confidence dropped

Without governance: Account could be locked automatically on a marginal or biased signal; dispute becomes reputational/legal incident.

Prevented failure mode: unowned reliance becomes action

Illustrative exposure: customer harm + trust loss

Note: exposure figures are illustrative; the structural claim is that the system intercepts and governs at the moment before consequence.

Evidence & Integrity

5. Evidence minimums and signals

- model_score
- model_version
- prompt_provenance
- data_lineage
- tool_call_trace
- confidence
- autonomy_threshold

6. Portable artefacts included in this pack

evidence_pack.json	Top-level decision record and links.
authority_contract.json	Who can authorise what, with what bounds.
policy_trace.ndjson	Versioned doctrine checks evaluated.
audit_export.ndjson	Runtime event log for audit ingestion.
provenance_chain.json	Illustrative model/prompt/tool-call chain integrity.
manifest.json + signature	Tamper-evident integrity for this entire bundle.

7. Integrity verification

Run **verify_manifest.py** in the root folder to recompute SHA-256 hashes and verify the signature (when cryptography or OpenSSL is available).

Expected output: **PASS** (no files changed).

This proof object is designed to be inspected without access to the originating systems.

Contact: founder@judgementspine.com | <https://judgementspine.com/>