SDL

STRUCTURAL DESIGN LABS

Insurable AI: Five Criteria for Verifiable Governance

Introduction

High-stakes AI deployment requires provable governance, not trust statements. Insurability requires measurable risk reduction through deterministic controls and verifiable evidence. This framework outlines five criteria that make AI systems underwritable, policy enforcement design, auditable decisions, persistent policy application, transparent enforcement, and continuous validation. These criteria are operationalised through SIR, a deterministic pre-inference gate that enforces policy before the model sees the input and emits signed, offline-verifiable audits. SIR is model agnostic and supports updatable domain packs, portable, versioned policy suites that can be refreshed independently to address evolving risks and regulations.

Five Criteria

1. Policy Enforcement Design

Policy must be enforced as core logic, not add-on controls.

Deterministic rules applied before inference create consistent outcomes. Add-on controls create unreliable results.

Validation: SIR demonstrates this through its rules-only enforcement layer.

Example: A prompt attempting unauthorised data access is blocked before reaching the model, with the decision fully auditable.

2. Auditable Decisions

All critical decisions must be observable, documented, and reconstructable.

SIR generates signed per-run certificates and an immutable audit ledger (ITGL) for full traceability.

Example: Every block or pass produces a certificate that reconstructs the exact policy violation or approval reason.

3. Persistent Policy Application

System behaviour must maintain policy application across contexts and changes.

SIR binds the policy hash and safety fingerprint to prove configuration integrity.

Example: A policy update is fingerprinted and verified. Any drift is detectable before operational impact.

4. Transparent Enforcement

Stakeholders must have visibility into decision-making and policy triggers.

SIR provides explainable decisions, pass or block, with full policy lineage documented in signed certificates.

Example: A compliance officer can read the certificate to see why a prompt was blocked, including the triggering rule.

5. Continuous Validation

Structured testing must confirm enforcement stability under normal and stress conditions.

SIR delivers offline-verifiable audits and benchmarks that demonstrate consistent policy enforcement outcomes.

Example: Benchmarks run across models show zero leaks on policy violating prompts, with proofs attached.

Insurance Underwriting Logic

Underwriters require quantifiable risk assessment, measurable metrics, documented failure patterns, and stress-tested resilience. SIR provides these through deterministic enforcement and verifiable evidence of containment, reducing measurable exposure and enabling more favourable underwriting terms.

Validation Evidence

Validation has been demonstrated across multiple model providers with consistent policy enforcement outcomes under structured testing. Key findings include platform agnostic enforcement, configuration integrity under change, and measurable risk reduction.

Sector-Specific Risk Reduction

Healthcare: patient safety through auditable reasoning pathways, regulatory compliance via persistent enforcement.

Financial services: fraud prevention with reduced false positives, operational risk control via transparent decisions.

Public infrastructure: decision transparency, accountability frameworks, operational continuity.

Implementation

SIR demonstrates complete implementation across all five criteria today. It is MIT open and runs locally in minutes with no vendor dependencies. Next steps are controlled pilots with sector partners to produce formal evidence packages for underwriting and compliance.

Conclusion

Insurable AI bridges theoretical safety and operational risk management through measurable governance criteria. By demonstrating quantifiable risk reduction, persistent policy application, and operational transparency, the framework enables commercially underwritten AI deployment in highrisk environments. This is measurable governance producing actuarial risk reduction.

Test SIR yourself on the live proof page: https://www.structuraldesignlabs.com/proof

Run offline verification:

```
curl -s https://raw.githubusercontent.com/SDL-HQ/sir-firewall/main/proofs/latest-audit.json | python3 tools/verify_certificate.py
```

Expected output: OK: Certificate signature valid and payload_hash matches.

Keywords: actuarial risk assessment, deterministic enforcement, measurable governance, insurance underwriting, operational transparency, persistent policy application, commercial risk reduction.