

Mapping AI Governance Risks to SFCL Safeguards

Bias & Fairness

SFCL Safeguard: SFCL enforces epistemic discipline. Every claim must be attributed, traceable, and scored for reliability.

Buyer Value: This helps organizations demonstrate fairness in decision-making and reduces legal exposure.

Over-reliance & De-skilling

SFCL Safeguard: Protective Modes (Containment, Solemn) require human judgment at checkpoints. **Buyer Value:** Prevents blind automation and maintains staff expertise.

Accountability

SFCL Safeguard: Claim Graphs & Attestations tie every fact to a source, role, and provenance. **Buyer Value:** Creates clear audit trails for 'who said what, when,' supporting legal defensibility and organizational trust.

Stakeholder Trust

SFCL Safeguard: Tone protocols (Tenderness, Dignity, Solemn) ensure safety, dignity, and clarity in delivery. **Buyer Value:** Builds trust with staff, clients, and funders, improving adoption.

Cost & Resource Management

SFCL Safeguard: Deterministic retrieval discipline, memory budgets, and eviction priorities prevent wasted compute cycles and hallucinations.

Buyer Value: Supports efficiency and sustainability in AI deployments.

Transparency / Black Box

SFCL Safeguard: Structured logging (§2.0.2.6) outputs a visible epistemic surface for each turn. **Buyer Value:** Allows decisions to be explained in plain, traceable terms to clients, regulators, or auditors.

Accuracy & Drift

SFCL Safeguard: Progressive Retrieval Escalation (PRE), MemoryMaintenance jobs, and recency decay on reliability scores keep outputs aligned with current facts.

Buyer Value: Protects against outdated or degraded model performance, ensuring consistent reliability.

