

Enterprise AI Trust, Risk and Compliance (AI TRiSM) Framework

A Comprehensive Enterprise Governance Framework for AI Trust, Risk, Security, Compliance, Model Assurance, Human Oversight, and Operational Resilience

Sanjeeve Kumar Gajadi*

*(HCL Technologies, Bengaluru, Karnataka, India)

Email: sanjeevekumar.g@hcltech.com

Abstract:

Artificial Intelligence is becoming embedded in enterprise decisions, customer interactions, software engineering, operations, and autonomous workflows. As adoption expands, organizations must manage not only model performance but also trust, risk, security, compliance, explainability, privacy, resilience, and accountability. This paper proposes an Enterprise AI Trust, Risk and Compliance (AI TRiSM) Framework that integrates enterprise architecture, AI governance, model risk management, cybersecurity, privacy, Responsible AI, legal compliance, observability, and lifecycle controls. The framework establishes governance domains for AI inventory, risk classification, model assurance, data and prompt governance, identity, security, human oversight, monitoring, incident response, regulatory evidence, and recertification. It supports predictive AI, Generative AI, Retrieval-Augmented Generation, and Agentic AI across hybrid and multi-cloud environments. The proposed approach enables enterprises to scale AI while maintaining measurable trust, bounded risk, defensible compliance, and clear accountability.

Keywords — AI TRiSM, AI Trust, AI Risk Management, AI Compliance, Responsible AI, Model Risk, AI Security, Generative AI, Agentic AI, Enterprise Governance

I. INTRODUCTION

AI systems increasingly influence consequential enterprise outcomes. Trust therefore depends on more than accuracy; organizations must understand how systems are governed, secured, monitored, explained, and controlled.

AI TRiSM provides an integrated management approach for trust, risk, and security. This paper extends that concept into an enterprise operating framework with explicit governance domains and lifecycle controls.

II. TRUST OBJECTIVES

Trust should be measurable through reliability, transparency, explainability, privacy, security, fairness, accountability, and consistent business outcomes.

Different use cases require different trust thresholds. A productivity assistant and a regulated decision service should not be governed identically.

III. AI INVENTORY AND CLASSIFICATION

Organizations should maintain an inventory of models, agents, AI-enabled applications, RAG services, third-party copilots, and embedded AI capabilities.

Each item should record owner, provider, purpose, data, users, autonomy, jurisdiction, criticality, and risk tier to support portfolio-level governance.

IV. RISK CLASSIFICATION

Risk classification should consider data sensitivity, decision impact, autonomy, external exposure, reversibility, regulatory scope, user population, and safety consequences.

Controls, testing, approval, monitoring, and recertification frequency should scale according to the assigned risk tier.

V. MODEL ASSURANCE

Model assurance should validate suitability, performance, robustness, limitations, bias, explainability, and security before production use.

Material changes to models, versions, prompts, features, or retrieval sources should trigger regression evaluation and updated approval where required.

VI. DATA, PROMPT, AND KNOWLEDGE GOVERNANCE

Training data, enterprise data, prompts, embeddings, vector stores, and knowledge sources should have clear ownership, provenance, quality, access rules, and retention controls.

RAG systems should preserve source permissions and validate retrieved content because manipulated or stale information can change model behavior.

VII. AI SECURITY

Security controls should address prompt injection, model theft, malicious inputs, data leakage, poisoned knowledge, insecure plugins, agent impersonation, and tool abuse.

Zero Trust, least privilege, workload identity, secure gateways, sandboxing, secret management, content filtering, and continuous threat detection are foundational.

VIII. HUMAN OVERSIGHT AND ACCOUNTABILITY

High-impact use cases should define when humans approve, review, override, or escalate AI outputs. Oversight should be meaningful rather than symbolic.

Named business owners remain accountable for outcomes even when AI performs significant work.

IX. COMPLIANCE MANAGEMENT

AI compliance should map applicable laws, sector rules, privacy obligations, contractual restrictions, internal policies, and documentation requirements to each use case.

Evidence should include risk assessments, approvals, evaluation results, access reviews, audit logs, incident records, and recertification decisions.

X. OBSERVABILITY AND CONTINUOUS CONTROL

Runtime telemetry should capture model version, prompts, retrieved sources, decisions, tool calls, policy checks, user feedback, errors, cost, and outcome quality.

Continuous control monitoring can identify drift, abnormal usage, policy violations, rising risk, and unsupported models before they create material incidents.

XI. INCIDENT RESPONSE

AI incident playbooks should cover sensitive-data disclosure, unsafe output, agent misuse, prompt injection, compromised models, poisoned sources, and regulatory breaches.

Response actions may include disabling tools, revoking credentials, switching models, blocking data sources, rolling back versions, or suspending the AI service.

XII. THIRD-PARTY AND SUPPLY-CHAIN RISK

Enterprises should evaluate AI providers, model licenses, data handling terms, hosting regions, dependency vulnerabilities, update practices, and contractual obligations.

Third-party changes should be monitored because provider updates can alter model behavior, privacy posture, or security risk.

XIII. AI GOVERNANCE OPERATING MODEL

Governance should involve business owners, enterprise architecture, security, model risk, data governance, legal, privacy, compliance, operations, and AI governance functions.

Decision rights should specify who approves use cases, models, data access, production release, major changes, and retirement.

XIV. KPIS AND KRIS

KPIs can include percentage of inventoried AI assets, approved models, evaluated high-risk use cases, current recertification, and monitoring coverage.

KRIs can include policy violations, high-risk incidents, unsupported models, unresolved audit findings, excessive privileges, and overdue recertification.

XV. IMPLEMENTATION ROADMAP

Organizations can begin with inventory, risk taxonomy, minimum control standards, and a small governance council. Higher-risk use cases should receive priority remediation.

Automation can then extend through policy-as-code, AI gateways, centralized telemetry, evaluation pipelines, and integrated compliance evidence.

XVI. CONCLUSION

Enterprise AI trust requires coordinated management of risk, security, compliance, transparency, and accountability across the entire lifecycle.

The proposed AI TRiSM framework provides a practical operating model for scaling predictive, generative, and agentic AI while maintaining defensible control and enterprise trust.

REFERENCES

- [1] ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system.
- [2] ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection — Information security management systems.
- [3] ISO 31000:2018, Risk management — Guidelines.
- [4] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), 2023.
- [5] National Institute of Standards and Technology, AI RMF Generative Artificial Intelligence Profile, 2024.
- [6] National Institute of Standards and Technology, SP 800-207: Zero Trust Architecture, 2020.
- [7] The Open Group, TOGAF Standard, 10th Edition, 2022.
- [8] ISACA, COBIT 2019 Framework: Governance and Management Objectives, 2019.
- [9] SAP SE, SAP Business Technology Platform Documentation.
- [10] SAP SE, SAP AI Core and SAP AI Launchpad Documentation.
- [11] SAP SE, SAP Integration Suite Documentation.
- [12] SAP SE, SAP HANA Cloud Documentation.
- [13] SAP SE, SAP Cloud Identity Services Documentation.
- [14] SAP SE, SAP Cloud ALM Documentation.

- [15] SAP SE, SAP Business Data Cloud and SAP Datasphere Documentation.
- [16] OWASP Foundation, OWASP Top 10 for Large Language Model Applications.
- [17] MITRE, ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems.
- [18] OECD, Recommendation of the Council on Artificial Intelligence.

AUTHOR BIOGRAPHY

Sanjeeve Kumar Gajadi is an SAP Enterprise Architect and Lead Consultant with extensive experience in enterprise architecture, SAP BTP, cloud, security, governance, data, and AI-enabled transformation. His professional interests include AI TRiSM, Responsible AI, enterprise AI security, model risk, and agentic AI governance.

ACKNOWLEDGMENT

The author acknowledges published guidance from enterprise architecture practitioners, standards bodies, AI governance communities, security researchers, and SAP documentation teams.