

Enterprise AI Supply Chain Security and Model Provenance Framework

A Comprehensive Enterprise Framework for Securing AI Models, Data, Components, Providers, Artifacts, Dependencies, and End-to-End Provenance

Sanjeeve Kumar Gajadi*

*(HCL Technologies, Bengaluru, Karnataka, India)

Email: sanjeevekumar.g@hcltech.com

Abstract:

Enterprise AI solutions depend on complex supply chains that include foundation models, open-source libraries, datasets, embeddings, model hubs, APIs, containers, prompts, retrieval components, plugins, agent tools, and third-party services. Weaknesses or tampering anywhere in this chain can introduce security, integrity, licensing, privacy, and operational risks. This paper proposes an Enterprise AI Supply Chain Security and Model Provenance Framework that combines asset inventory, trusted sourcing, software and model bills of materials, provenance metadata, cryptographic integrity, vulnerability management, dependency analysis, provider assurance, artifact signing, environment controls, and continuous monitoring. The framework defines controls across acquisition, development, training, fine-tuning, deployment, update, and retirement. It also addresses model lineage, dataset lineage, prompt provenance, external tools, agent dependencies, and evidence for audit and incident response. The proposed approach enables enterprises to verify what AI components are used, where they originated, how they changed, and whether they remain trusted throughout the lifecycle.

Keywords — AI Supply Chain Security, Model Provenance, AI Security, Model Lineage, SBOM, AI BOM, Software Supply Chain, Generative AI, Agentic AI, Enterprise Governance

I. INTRODUCTION

AI supply chains combine software, models, data, services, and external providers, creating attack surfaces beyond conventional application dependencies.

Enterprises need provenance to determine the origin, integrity, ownership, and transformation history of each AI component.

II. SUPPLY CHAIN ASSET INVENTORY

Inventory should include models, model versions, datasets, libraries, containers, prompts, embeddings, vector stores, APIs, plugins, tools, and external services.

Each asset should have an owner, source, version, license, integrity status, risk tier, and lifecycle state.

III. TRUSTED SOURCING

Acquisition policies should define approved model hubs, repositories, providers, package sources, and validation requirements.

Unverified models or datasets should not enter production pipelines without review of origin, licensing, integrity, and security.

IV. MODEL AND COMPONENT BILLS OF MATERIALS

An AI bill of materials can extend software bills of materials to include model, dataset, prompt, embedding, tool, and provider dependencies.

Machine-readable manifests improve vulnerability response, impact analysis, and auditability.

V. PROVENANCE AND LINEAGE

Provenance should record creation source, transformations, training or fine-tuning inputs,

evaluation results, approvals, signatures, and deployment targets.

Lineage must remain intact across model promotion, quantization, adaptation, prompt updates, retrieval changes, and provider version changes.

VI. INTEGRITY AND ARTIFACT SIGNING

Models, containers, configuration, prompts, and deployment packages should use checksums, signatures, attestations, or equivalent integrity controls.

Deployment pipelines should verify artifacts before execution and block untrusted or modified components.

VII. DATASET AND TRAINING DATA SECURITY

Training and evaluation data should be scanned for sensitive information, poisoning indicators, licensing restrictions, and integrity issues.

Dataset provenance should support reproducibility and allow affected models to be identified when source data is later found to be problematic.

VIII. THIRD-PARTY PROVIDER RISK

Providers should be assessed for secure development, model update practices, incident response, subcontractors, data handling, availability, and provenance transparency.

Contracts should address notification of material model changes, security incidents, and changes in data-use practices.

IX. OPEN-SOURCE AND DEPENDENCY SECURITY

Libraries and frameworks should be monitored for vulnerabilities, malicious packages, abandoned dependencies, and incompatible licenses.

Dependency pinning, trusted registries, code scanning, and controlled build environments reduce supply-chain exposure.

X. AGENT TOOLS AND PLUGIN SUPPLY CHAIN

Agentic AI extends the supply chain to tools, plugins, APIs, and external capabilities that can execute actions.

Tool provenance, ownership, permissions, versioning, and security posture should be verified before agents are allowed to use them.

XI. CONTINUOUS MONITORING AND RESPONSE

Continuous monitoring should identify new vulnerabilities, provider changes, integrity failures, unexpected model updates, and dependency drift.

Impact analysis should use bills of materials and lineage to quickly determine which applications and business processes are affected.

XII. GOVERNANCE AND AUDIT EVIDENCE

Governance should define approval criteria, exception handling, retention of provenance evidence, and risk ownership.

Auditors should be able to reconstruct the chain from source model and data through build, evaluation, approval, deployment, and current runtime version.

XIII. CONCLUSION

The Enterprise AI Supply Chain Security and Model Provenance Framework provides a structured method for managing trust across the full AI dependency chain.

By combining inventory, provenance, integrity verification, provider assurance, and continuous monitoring, enterprises can reduce hidden AI supply-chain risk and improve incident response and auditability.

ACKNOWLEDGMENT

The author acknowledges the contributions of enterprise architecture practitioners, AI governance specialists, security professionals, standards bodies, and technology researchers whose published work informed the concepts synthesized in this paper.

REFERENCES

- [1] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), 2023.
- [2] National Institute of Standards and Technology, AI RMF Generative Artificial Intelligence Profile, 2024.
- [3] ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system.
- [4] ISO 31000:2018, Risk management — Guidelines.

- [5] ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection — Information security management systems.
- [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] OWASP Foundation, OWASP Top 10 for Large Language Model Applications.
- [10] MITRE, ATLAS: Adversarial Threat Landscape for Artificial-Intelligence Systems.
- [11] OECD, Recommendation of the Council on Artificial Intelligence.
- [12] IEEE, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems.