World Digital Technology Academy (WDTA)

Large Language Model Security Requirements for Supply Chain

World Digital Technology Academy Standard

WDTA AI-STR-03

Edition: 2024-09

© WDTA 2024 – All rights reserved.

The World Digital Technology Standard WDTA AI-STR-03 is designated as a WDTA norm. This document is the property of the World Digital Technology Academy (WDTA) and is protected by international copyright laws. Any use of this document, including reproduction, modification, distribution, or re-publication, without the prior written permission of WDTA, is prohibited. WDTA is not liable for any errors or omissions in this document.

Discover more WDTA standards and related publications at **https://wdtacademy.org/.**

Version History*

Standard ID	Version	Date	Changes
WDTA AI-STR-03	1.0	2024-09	Initial Release

Foreword

As the digital age advances, the integration of artificial intelligence, particularly large language models (LLMs), has become a cornerstone of modern technological ecosystems. These models are now pivotal in shaping industries, driving innovation, and transforming the way we interact with technology. However, with this rapid integration comes an array of security challenges that must be addressed to ensure these powerful tools' safe and responsible deployment.

The World Digital Technology Academy (WDTA) has long been at the forefront of setting global standards for digital technology and innovation. Our commitment to fostering a secure and inclusive digital world is reflected in the rigorous development of standards that guide the deployment and management of cutting-edge technologies. The AI STR (Security, Trust, Responsibility) series, to which this document belongs, is a collection of standards designed to ensure that AI technologies are innovative, secure, trustworthy, and ethically managed. These standards provide comprehensive frameworks for addressing the complex challenges associated with AI deployment, focusing on critical aspects like safety, integrity, and responsible use.

As the WDTA AI-STR-03 standard, the "Large Language Model Security Requirements for Supply Chain" outlines comprehensive measures for managing security risks across the supply chain of large language models. This standard covers the entire lifecycle of these models, from development through deployment, ensuring that each phase is rigorously scrutinized for potential vulnerabilities. By adhering to these guidelines, organizations can effectively protect their AI-driven operations from emerging threats and contribute to a more secure digital ecosystem.

We extend our gratitude to the experts and contributors who have worked diligently to develop this standard. Their expertise and commitment to excellence ensure that WDTA continues to be a leader in setting the benchmark for digital security. We encourage all stakeholders in the AI supply chain to adopt these guidelines, helping to build a future where technological advancement goes hand in hand with security and ethical responsibility.

Executive Chairman of WDTA

Acknowledgments

Co-Chair of WDTA AI STR Working Group

Ken Huang *(CSA GCR)*

Josiah Burke *(Anthorphic)*

Lead Authors

Jiashui Wang *(Ant Group)*

Weiqiang Wang *(Ant Group)*

Long Liu *(Ant Group)*

Yuhao Jiang *(Ant Group)*

Ken Huang *(CSA GCR)*

Anyu Wang *(CSA GCR)*

Zheng Song *(Ant Group)*

Jiawei Tang *(Ant Group)*

Yin Wang *(Ant Group)*

Zhihui Jiang *(Ant Group)*

Liang Zheng *（Ant Group）*

Cong Zhu *（Ant Group）*

Qing Luo *（Ant Group）*

Shiwen Cui *（Ant Group）*

Miao Chen *(Zhongguancun Laboratory)*

Tianyu Cui *(Zhongguancun Laboratory)*

Reviewers

Lars Ruddigkeit *（Microsoft）*

Anton Chuvakin *（Google）*

Apostol Vassilev *(NIST)*

Dongchen Ma *（Tencent Cloud)*

Chenfu Bao *（Baidu)*

Feng Luo *(Shenzhen National Financial Technology Testing Center)*

Haoshuo Wang *(China Mobile Cloud Centre)*

Melan XU *(World Digital Technology Academy)*

Tal Shapira *(Reco AI)*

Dr. Cari Miller *(Center for Inclusive Change)*

Govindaraj Palanisamy *(Global Payments Inc.)*

Krystal (A) Jackson *(Frontier Model Forum)*

Swapnil Modak *(Meta)*

Heather Frase *(verAItech)*

Vishwas Manral *(Precize Inc)*

Patricia Thaine *(Private AI)*

Liming Zhang *(Comcast)*

Vaibhav Malik *(Cloudflare)*

Asha Hemrajani *(Nanyang Technological University)*

Ron F. Del Rosario *(SAP ISBN)*

Madhavi Najana *(Federal Home Loan Bank of Cincinnati)*

Gaurav Puri *(META)*

Bhuvaneswari Selvadurai *(Northwestern Mutual)*

Dan Stocker *(Coalfire)*

Matteo Meucci *(IMQ MINDED SECURITY)*

Qiang Zhang *(Coalfire)*

Joshuaanaguiar *(Cohere)*

Daemon Behr *(Nutanix)*

Table of Contents

1 Scope

This document presents the framework of supply chain security protection for large language models (LLMs), proposes requirements for managing supply chain security risks and supply activities involved in the development, operation, and maintenance (O&M) of LLMs, and provides relevant information such as common supply chain security risks and typical security cases.

This document can guide suppliers and consumers in the supply chain in carrying out security risk assessment and managing supply activities. It can also serve as a foundation for third-party organizations conducting supply chain security tests and assessments for regulatory authorities.

2 Normative References

The following documents constitute essential provisions of this document through normative references in the text. For dated reference documents, only the version corresponding to the date applies to this document; for undated reference documents, the latest version (including all amendments) applies to this document.

ISO 28001	Security management systems for the supply chain – Best practices for implementing supply chain security, assessments and plans – Requirements and guidance
ISO/IEC 27036-2	Information technology – Security techniques – Information security for supplier relationships – Part 2: Requirement
ISO/IEC 27036-3	Information technology – Security techniques – Information security for supplier relationships – Part 3: Guidelines for information and communication technology supply chain security

NIST 800-161	Supply Chain Risk Management Practices for Federal Information Systems and Organizations
NIST AI RMF 1.0	Artificial Intelligence Risk Management Framework
ISO 42001/IEC	ISO/IEC 42001 is an international standard that specifies requirements for establishing, implementing, maintaining, and continually improving an Artificial Intelligence Management System (AIMS) within organizations. It is designed for entities providing or utilizing AI-based products or services, ensuring responsible development and use of AI systems.
ISO/IEC 5338	Information technology — Artificial intelligence — AI system life cycle processes. It is an international standard that defines a set of processes and associated concepts for describing the life cycle of AI systems. This standard is particularly focused on AI systems based on machine learning and heuristic methods. It builds on existing standards like ISO/IEC/IEEE 15288 and ISO/IEC/IEEE 12207, incorporating AI-specific processes from ISO/IEC 22989 and ISO/IEC 230531. The goal is to provide a comprehensive framework for the definition, control, management, execution, and improvement of AI systems throughout their life cycle.
GB/T 36637-2018	Information security technology - Guidelines for the information and communication technology supply chain risk management
GB/T 43698-2024	Cybersecurity technology - Security requirements for software supply chain

GB/T 24420-2009	Supply chain risk management guideline
GB/T 32921-2016	Information security technology - Security criterion on supplier conduct of information technology products

3 Terms and Definitions

The following terms and definitions apply to this document.

3.1 Artificial Intelligence

Artificial intelligence (AI) is a multifaceted field within computer science focused on creating systems that can perform tasks typically requiring human intelligence. An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

3.2 Large Language Model

Large-scale, pre-trained and fine-tuned AI models that can understand instructions and generate outputs across multiple modalities, including but not limited to human languages, program codes, images, and audio, based on large amounts of data.

3.3 Supplier

An organization or individual develops, produces, augments, adapts, finetunes, provides, and/or deploys software products or services.

3.4 Software supply chain

A network chain system that delivers software products or services from suppliers to consumers through resources and processes based on the relationship.

3.5 Open source community

An organization and operation mode for developing and maintaining open-source code.

3.6 Third-party component

Independent or callable software components developed by software development organizations or personnel other than suppliers and consumers usually consist of binary or source code program files.

3.7 Machine Learning Platform

An integrated environment that provides support and tools for developing, training, and deploying machine learning models.

3.8 Large Language Model Inference Framework

An integrated environment dedicated to deployment and the performance of model inference.

3.9 Large Language Model Application Framework

Application development framework based on LLMs.

3.10 Distributed Computing Framework

A framework for processing large amounts of data in parallel on multiple computers.

3.11 Machine Learning Bill of Materials

A list of standardized model cards, models, datasets, data cards, system cards, and other materials involved in building an LLM model.

4 Overview of Supply Chain Security Protection for LLMs

The main objective of this document is to identify, evaluate, and manage the supply chain security risks in the LLM system life cycle. LLMs may be used in services, operated in whole or in part by third parties, or as products, received from third parties, but not operated by them.

4.1 Supply Chain Security for LLMs

The supply chain usually covers the procurement, development, integration, and other phases of software and hardware products. It involves producers, suppliers, system integrators, service providers, other entities, and soft environments such as technology, law, and strategy. Unlike traditional supply chains, the LLM supply chain covers the entire life cycle of the LLM, including model and training data acquisition, training data preparation, model training, fine-tuning, deployment, operations and maintenance (O&M), and other stages.

Supply chain security management for LLMs involves two types of security requirements. One is general security requirements throughout the life cycle, called Supply Chain Security Management for LLMs, such as requirements for procedures, organizations, personnel, and information systems related to supply chain security management. The other is security requirements related to the system structure of LLMs, called supply chain security requirements for LLMs, which include requirements for the network layer, system layer, platform and application layer, model layer, and data layer.

4.2 Objectives of Supply Chain Security Management for LLMs

a) Integrity: Ensure that the product and its systems, components, frameworks, models, data, and used tools are protected against implantation, tampering, or unauthorized replacement throughout the entire life cycle of LLM products. This involves the implementation of rigorous controls and continuous monitoring at every stage of the supply chain. Including, addressing common vulnerabilities in middleware security to prevent unauthorized access, safeguarding against the risk of poisoning training data used by engineers, and enforcing a zero-trust architecture to mitigate internal threats. By maintaining the integrity of every stage, from data

acquisition to supplier deployment, consumers using LLMs can ensure that the LLM products remain secure and trustworthy.

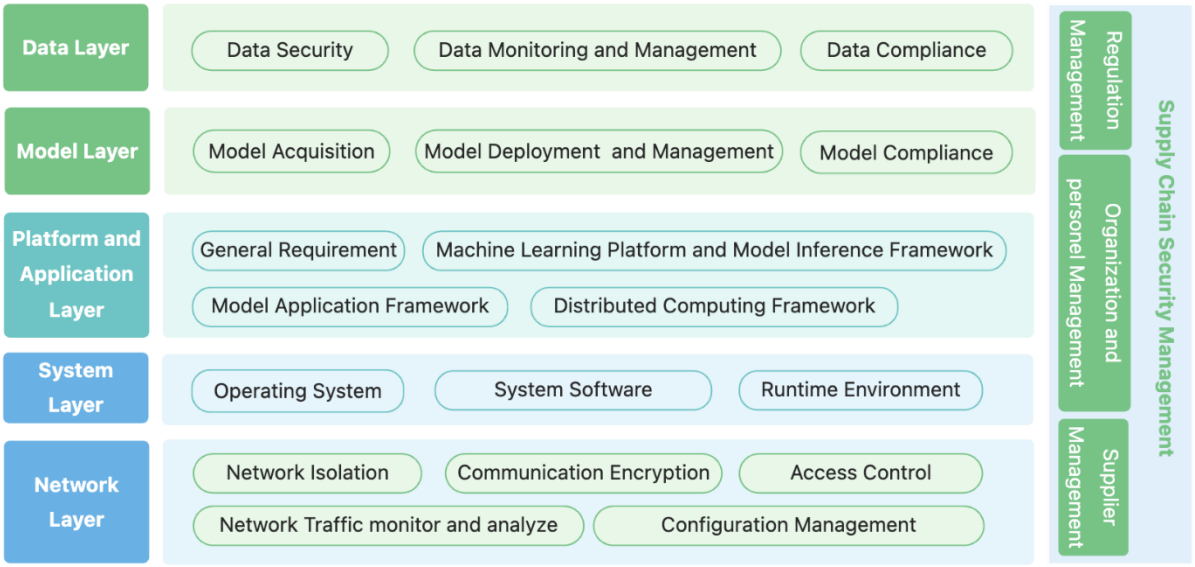
b) Availability: Ensure the supply chain's availability for consumers. Suppliers must supply materials by agreements concluded and signed with consumers without interruption by human or natural factors. Additionally, they should ensure that the supply can be predictably recovered to an acceptable state under certain conditions if it were to partially fail.

c) Confidentiality: Ensure that information transmitted along the supply chain is not disclosed to unauthorized persons, including information about the consumers themselves.

d) Controllability: Guarantee consumers' meaningful control over the supply chain. Ensure consumers have an understanding of information in all phases of the supply chain, transparency and credibility of suppliers/service providers at all levels, management of data flow, and traceability of the supply chain.

e) Reliability: Ensure the security, high availability, and Disaster tolerance of LLM products and relevant systems, components, frameworks, models, and data.

f) Visibility: Ensure supply chain steps, changes, updates, and deletions with every step of the change are trackable, have clear ownership, and can be traced back as needed. For example, if a model is updated with new training data, the training data and the model before and after the training should be documented and traceable.



5 Supply Chain Security Management for LLMs

5.1 Regulation Management

a) Formulate supply chain security management policies and procedures, including but not limited to risk management regulations, processes, and mechanisms for software supply phases such as procurement, delivery, and O&M.

b) Formulate regulations for continuous risk monitoring, risk assessment, and incident response for the software supply chain. These regulations can contain emergency response procedures, operation halting procedures, system recovery procedures, timely notices to backward and forward supply chain members, and other content.

c) Conduct supply chain risk assessments regularly, formulate plans, and take measures to eliminate or reduce risks. These assessments should cover all stages of the LLM supply chain, identifying potential vulnerabilities like third-party risks or data integrity issues. Use the results to develop targeted mitigation strategies and update security policies, ensuring continuous protection against emerging threats.

d) Establish governance frameworks for LLM development that enforce compliance with security standards and industry best practices throughout the supply chain.

5.2 Organization and Personnel Management

a) Propose security requirements for supply chain security management personnel, including but not limited to administrators, architectural engineers, AI engineers, Data Scientists, ordinary employees, and third-party personnel.

b) Formulate security training plans and carry out regular supply chain security training. The training should include, without being limited to, software asset identification and analysis, integrity guarantee, guardrails, and software vulnerability and backdoor analysis.

c) Conduct security awareness and skill training for all internal employees with additional specialized training for those involved in supply chain and security management such as

procurement, information system development and management, and product O&M based on corresponding security requirements.

5.3 Supplier Management

a) Develop supplier selection strategies and regulations and conduct security assessments of suppliers for self-developed software, customized software, off-the-shelf software, and different aspects of other software as well, including but not limited to the background, capability, qualification, and continuous and secure provision of products or services. Maintain an inventory of all AI solutions and assets, including but not limited to AI source suppliers, supply chain members, model types, internal owners, last security reviews, etc.

b) Suppliers must ensure and attest to the authenticity, accuracy, and integrity of the information transmitted along the software supply chain and take measures to avoid tampering and leakage.

c) Require suppliers to cooperate in security monitoring and inspection of the software supply chain, including through periodic or on-demand independent audits.

6 Supply Chain Security Requirements for LLMs

6.1 Network Layer

a) Segment networks to isolate critical data and systems from external and internal networks: Implement network segmentation to reduce attack surfaces, ensuring that sensitive information is inaccessible from unauthorized network zones/unauthorized resources, aligning with the Zero Trust principle of least privilege.

b) Enforce secure encryption for all network communications: Apply robust encryption protocols for data in transit to maintain confidentiality and integrity, ensuring no communication is trusted by default, even within internal networks.

c) Implement strict access controls with continuous monitoring: Deploy security access controls that enforce the least privilege for accessing critical information and services, with

continuous logging and monitoring of access events to detect and respond to potential threats in real time.

d) Continuously monitor and analyze network traffic for anomalies: Utilize advanced monitoring solutions to scrutinize network activities continuously, swiftly identifying and mitigating abnormal behaviors per the MITRE ATT&CK and Atlas framework.

e) Regularly audit and maintain network security configurations: Conduct continuous security audits and proactive maintenance on network devices such as routers, switches, and firewalls, ensuring that patches and updates are promptly applied to mitigate vulnerabilities, in line with Zero Trust's emphasis on ongoing security posture assessment.

6.2 System Layer Security Requirements

The system layer mainly targets the supply chain security requirements for the underlying LLM operating system, system software, and runtime environment.

6.2.1 Operating System Security Requirements

a) Update the operating system regularly and install security patches promptly to prevent system security vulnerability attacks.

b) Establish a system update mechanism and verification process to ensure that security patches are installed promptly and accurately.

c) Implement access control measures to manage the access of users and programs to operating system functions, including but not limited to user authentication, authorization, and auditing.

d) Establish operating system security monitoring and protection capabilities to monitor and prevent suspicious activities or security incidents on time.

6.2.2 System Software Security Requirements

a) Install system software from trusted sources and verify its integrity and authenticity.

b) Update system software regularly and install security patches in time to prevent security vulnerability attacks.

c) Digital signature technology should be used to verify software security.

d) Configure system software according to security best practices to avoid security risks from default configurations.

e) Regular audits and compliance checks of device security configuration are conducted.

6.2.3 Runtime Environment Security Requirements

a) Use virtualization or container technology to create a secure, independent, and isolated operating environment for each application to reduce possible security risks and impact.

b) Deploy comprehensive monitoring and anomaly detection systems, including but not limited to resource usage, performance metrics, and security events.

c) Record all critical runtime activity logs and conduct regular audits to prevent unauthorized access and tampering.

d) Audit the runtime environment regularly to identify and fix security configuration errors promptly.

e) The processing and storage of sensitive data shall be conducted in a trusted computing environment.

6.3 Platform and Application Layer Security Requirements

The platform and application layer includes machine learning frameworks and other third-party components and is the critical support environment for operating large language models.

6.3.1 General Security Requirements for Components

a) Manage third-party components strictly in terms of their source and version to ensure timely updates and proper security.

b) Conduct regular security checks on third-party components and upgrade them promptly to the latest secure versions.

c) Ensure imported third-party components undergo security reviews, including code audits and dependency analysis, to prevent the introduction of components with security risks.

d) Perform file parsing in a sandbox or similarly isolated environment to prevent security risks from potential memory corruption vulnerabilities in text parsing components.

e) Necessary licenses and authorizations should be obtained before using third-party components. Ensure that the components are used legally and in compliance with all relevant copyright and usage agreements.

f) Establish and maintain a software bill of materials regularly.

6.3.2 Machine Learning Platform and Model Inference Framework Security Requirements

Model inference frameworks need to be used during model deployment and runtime. At this stage, the model shall be considered as an executable program and attention shall be paid to code execution risks during model inference.

a) Conduct security analysis and checks on the model files before running the model.

b) Continuously validate model integrity of an inferencing model. Any auto-updates to model artifacts need to be tracked. Implement at least the two-person integrity rule to any model artifact updates to prevent unauthorized actions.

c) When using third-party models, carefully enable parameters that trust remote code, such as 'trust_remote_code' in transformers library, to reduce the risk of malicious code execution.

d) Use trusted model files to avoid security risks arising from the execution of malicious code.

6.3.3 Model Application Framework Security Requirements

a) Keys for invoking LLM interfaces shall not be stored in code.

b) Use guardrails and other detective controls in LLM applications to improve steerability and reduce risks such as prompt injection.

c) When using code interpreters or other code execution tools, employ secure isolation techniques such as containers or sandboxes.

d) File path checks should be performed to prevent path traversal vulnerabilities when using file processing tools.

e) Access control measures should be taken to prevent unauthorized operations when using database processing tools.

6.3.4 Security Requirements for Distributed Computing Frameworks

The primary risk for distributed computing frameworks comes from the need for permission checks between root nodes and child nodes in many distributed frameworks, allowing devices on the same network to connect directly to nodes and send commands.

a) Establish network isolation to prevent potential external attackers from accessing distributed computing nodes. Uses firewalls and intrusion detection systems to monitor and control traffic from child nodes. Decrypt data received from the root node and encrypt any data sent back to the root node.

b) Employ an authority verification mechanism between the root and child nodes to prevent malicious node connections or command execution. Use role-based access control to ensure that only authorized child nodes can communicate with the root node.

c.) Root Nodes ensure that the entire distributed framework adheres to cybersecurity standards such as NIST, CIS, ISO, and others. Child nodes comply with the root node's security policies and standards.

6.4 Model Layer Security Requirements

6.4.1 Model Acquisition Security Requirements

a) Obtain model files from trusted third parties and authorized model repositories.

b) Conduct integrity checks on model files obtained from third parties to ensure they have not been tampered with during storage and transmission.

c) Perform security checks, including pickle scanning, on model files obtained from third parties, to prevent the execution of malicious code or other security risks.

6.4.2 Model Deployment and Management Security Requirements

a) Deploy tools to monitor model behavior and promptly detect and respond to such behavior.

b) Establish and maintain a machine learning bill of materials(ML-BOM) regularly. The ML-BOM should document model architectures, version histories, training and fine-tuning datasets with their sources and preparation methods, pre-trained base models and their origins, custom algorithms or techniques used in model development, software libraries, and their versions, hardware specifications for training and inference environments, and model cards with performance metrics and intended use cases.

c) Implement a secure, version-controlled system for the ML-BOM, ensuring only authorized personnel can access and modify it while tracking changes over time to support auditing and enable rollback if necessary. Integrate the ML-BOM with existing development and deployment pipelines for automatic updates.

d) Utilize the ML-BOM to facilitate supply chain risk assessments, compliance audits, and incident response, enabling quick identification of affected components in case of a security event. Leverage the ML-BOM for enhanced security measures by using it to verify the integrity of model components during deployment, cross-referencing it with vulnerability databases to proactively identify potential risks, and employing it in conjunction with monitoring tools to detect unexpected changes in model behavior that might indicate a supply chain attack.

e) Regularly update the ML-BOM to reflect any changes in the model or its components, ensuring it remains a current and accurate representation of the LLM's supply chain.

6.4.3 Model Compliance Security Requirements

a) Ensure that model development and deployment processes comply with relevant regulations and standards.

b) Evaluate models using responsible AI principles, including harm and safety criteria, before advancing further development.

c) When utilizing third-party models, it is recommended to select models with detailed disclosures to prevent copyright infringement or legal issues. Clarify liabilities clauses for Open Source and closed vendor models.

6.5 Data Layer Security Requirements

Data layer security is the basis for ensuring the security of the large language model supply chain and involves the security of data collection, storage, processing, and transmission. This chapter does not cover the security issues of third-party model training data.

6.5.1 Data Security Requirements

a) Implement access control measures to protect data from unauthorized access.

b) Use role-based access control (RBAC) to restrict data access based on user roles and responsibilities.

c) Enforce the principle of least privilege, granting users only the minimum necessary access.

d) Encryption technology is used to ensure the security of data storage and transmission.

e) Conduct data consistency checks and integrity verification to ensure that the consistency and integrity of the data are not compromised throughout its life cycle.

f) Back up important data regularly and establish an effective disaster recovery plan to ensure that operations can be quickly restored in the event of data loss or system failure.

g) Manage lineage and traceability of all data used in Machine learning using data cards and data catalogs.

6.5.2 Data Compliance Security Requirements

a) In the training stage of the model, review data sources for security to ensure that all data comes from legitimate, reliable sources.

b) Obtain the necessary permissions and authorizations before using third-party data. Ensure the legality of data usage and compliance with all relevant copyright and usage agreements.

c) Desensitize or anonymize sensitive data identified to ensure that its use does not infringe on personal privacy or corporate secrets.

d) Define and publicly disclose data collection, usage, and storage policies. Establish clear roles for data compliance within the organization and assign specific responsibilities for each data collection and usage step.

e) Conduct compliance audits on the data used regularly to ensure compliance with legal and policy requirements.

f) Establish a monitoring mechanism to detect and correct possible compliance issues promptly.

6.5.3 Data Monitoring and Management

a) Data shall be rigorously validated and cleansed to exclude data points with incorrect format, bias, incompleteness, or apparent abnormalities before input into the model.

b) Introduce data cleaning tools and algorithms to identify and process potentially malicious information in unstructured data, such as embedded malicious code or links.

c) Identify outliers or unexpected data patterns in the dataset, including but not limited to using statistical methods and machine learning techniques.

d) Implement a dynamic detection mechanism to monitor data flows in real time and quickly respond to potential data contamination events.

e) Control and verify data sources strictly to ensure that all data sources are reliable and have a good security record.

f) Implement additional review and processing procedures for data collected from external or less trusted data sources.

g) Check the integrity of the data to ensure the data integrity has not been compromised during collection, transmission, or storage.

7 Summary

The WDTA AI-STR-03 standard presents a framework for managing security risks in the Large Language Model (LLM) supply chain. It addresses the unique challenges posed by the integration of AI technologies, particularly LLMs, into modern technological ecosystems. The standard covers the entire lifecycle of LLMs, from development and training to deployment and maintenance, providing detailed guidelines for each stage.

At its core, the standard emphasizes a multi-layered approach to security, encompassing network, system, platform and application, model, and data layers. It leverages key concepts such as the Machine Learning Bill of Materials (ML-BOM), Zero Trust Architecture, and continuous monitoring and auditing. These concepts are designed to ensure the integrity, availability, confidentiality, controllability, and reliability of LLM systems throughout their supply chain.

Model developers can leverage this standard document to enhance their ability to identify, evaluate, and manage supply chain security risks in LLM systems. The standard not only addresses technical aspects but also covers organizational and compliance requirements, reflecting the complex and interdisciplinary nature of LLM development and deployment. As AI technologies continue to evolve and become more integral to various industries, the WDTA AI-STR-03 standard provides a practical foundation for building secure, trustworthy, and ethically managed AI systems.