TrustIoT Framework for Industry 4.0

"AI Model Integrity and Confidentiality"

|  |  |
| --- | --- |
| Document Classification: | Internal |
| Document Ref. | *TrustIoT Framework for Industry 4.0* |
| Version: | *1* |
| Document Author: | *Jibran Saleem* |
| Document Owner: |  |

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Revision Author** | **Summary of Changes** |
|  |  |  |  |
|  |  |  |  |

**Distribution**

|  |  |
| --- | --- |
| **Name** | **Title** |
|  |  |
|  |  |
|  |  |

**Approval**

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Position** | **Signature** | **Date** |
|  |  |  |  |

Table of Contents

[1. Introduction 4](#_Toc190905223)

[2. Purpose 4](#_Toc190905224)

[3. Scope 4](#_Toc190905225)

[4. Policy Statement 4](#_Toc190905226)

[4.1. Secure Model Development and Training 4](#_Toc190905227)

[4.2. Secure Model Storage and Transmission 4](#_Toc190905228)

[4.3. Secure Model Deployment 4](#_Toc190905229)

[4.4. Access Control and Authorisation 5](#_Toc190905230)

[4.5. Monitoring and Integrity Checks 5](#_Toc190905231)

[5. Responsibilities 5](#_Toc190905232)

[6. Breaches of Policy 5](#_Toc190905233)

[7. Document Management 5](#_Toc190905234)

# Introduction

Artificial Intelligence (AI) and Machine Learning (ML) models are increasingly integrated into IoT systems to enable advanced analytics, automation, and decision-making capabilities. These models, often representing significant intellectual property and trained on sensitive data, require robust protection against unauthorised access, modification, or theft. This policy outlines the measures to ensure the integrity and confidentiality of AI models throughout their lifecycle, from development to deployment.

# Purpose

The purpose of this policy is to establish a framework for safeguarding the integrity and confidentiality of AI models used within the organisation's IoT infrastructure. This policy aims to:

* Protect AI models from unauthorised access, modification, and theft.
* Ensure the integrity and reliability of AI model outputs.
* Prevent the leakage or misuse of sensitive data used for training or inference.
* Maintain the organisation's competitive advantage by protecting its intellectual property.

# Scope

This policy applies to all AI and ML models developed, deployed, or utilised within the organisation's IoT environment, regardless of their specific purpose or application.

# Policy Statement

## Secure Model Development and Training

* **Secure Development Environment:** AI models shall be developed and trained in a secure environment with restricted access, utilising version control systems and secure coding practices.
* **Data Protection:** Data used for training AI models shall be classified based on its sensitivity and protected accordingly, using encryption and access controls.
* **Data Provenance:** The origin and lineage of training data shall be tracked and documented to ensure its authenticity and integrity.

## Secure Model Storage and Transmission

* **Encrypted Storage:** AI models shall be stored in encrypted format, both at rest and during transmission, using strong encryption algorithms and secure key management practices.
* **Access Controls:** Access to stored models shall be restricted to authorised personnel and systems, utilising role-based access control (RBAC) or similar mechanisms.
* **Secure Transmission Protocols:** Secure protocols, such as HTTPS or SFTP, shall be used for transmitting AI models between systems or environments.

## Secure Model Deployment

* **Protected Environments:** Deployed AI models shall be executed in secure environments, such as trusted execution environments (TEEs) or hardened containers, to protect against tampering and unauthorised access.
* **Integrity Verification:** Mechanisms shall be implemented to verify the integrity of deployed models and detect any unauthorised modifications.
* **Input Validation and Sanitisation:** Input data provided to AI models shall be validated and sanitised to prevent adversarial attacks or injection of malicious data.

## Access Control and Authorisation

* **Role-Based Access:** Access to AI models, training data, and model outputs shall be granted based on user roles and responsibilities within the organisation.
* **Need-to-Know Basis:** Access shall be granted on a need-to-know basis, limiting exposure to sensitive information.
* **Regular Reviews:** Access rights shall be periodically reviewed and updated to ensure they remain appropriate and aligned with current roles and responsibilities.

## Monitoring and Integrity Checks

* **Model Performance Monitoring:** The performance and accuracy of deployed AI models shall be continuously monitored to detect any anomalies or degradation that may indicate tampering or compromise.
* **Integrity Checks:** Periodic integrity checks shall be performed on stored and deployed models to verify their authenticity and detect any unauthorised modifications.
* **Logging and Auditing:** All access to and modifications of AI models shall be logged and audited for traceability and accountability.

# Responsibilities

* **Information Security Officer:** Responsible for overseeing the implementation and enforcement of this policy.
* **Data Scientists and AI Developers:** Responsible for developing, training, and deploying AI models in accordance with this policy.
* **IT Department:** Responsible for providing secure infrastructure and tools for AI model management and protection.
* **Model Owners:** Responsible for ensuring the ongoing security, integrity, and confidentiality of their respective AI models.

# Breaches of Policy

Non-compliance with this policy may result in disciplinary action, up to and including termination of employment or contractual relationships. Additionally, unauthorised access, modification, or theft of AI models may result in significant financial loss, reputational damage, and legal consequences for the organisation.

# Document Management

This document is valid as of [dd/mm/yyyy].

This document is reviewed periodically and at least annually to ensure compliance with the following prescribed criteria.

* Compliant with the Internet of Things (IoT) Security Framework for Industry 4.0.
* Legislative requirements defined by law, where appropriate.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

[Name 1]

Manager