TrustIoT Framework for Industry 4.0

"AI Accountability and Responsibility

|  |  |
| --- | --- |
| 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](#_Toc176334460)

[2. Purpose 4](#_Toc176334461)

[3. Scope 4](#_Toc176334462)

[4. Policy Statement 4](#_Toc176334463)

[4.1. Clear Responsibility and Accountability 4](#_Toc176334464)

[4.2. Documentation and Traceability 4](#_Toc176334465)

[4.3. Human Oversight and Control 4](#_Toc176334466)

[4.4. Impact Assessments and Risk Management 5](#_Toc176334467)

[4.5. Ethical Considerations 5](#_Toc176334468)

[5. Responsibilities 5](#_Toc176334469)

[6. Breaches of Policy 5](#_Toc176334470)

[7. Document Management 5](#_Toc176334471)

# Introduction

The increasing integration of Artificial Intelligence (AI) into Internet of Things (IoT) systems introduce complexities in decision-making and actions. It is imperative to establish clear lines of accountability and responsibility for AI-driven outcomes to ensure transparency, ethical conduct, and adherence to legal and regulatory requirements. This policy outlines the framework for establishing and maintaining accountability in the context of AI-powered IoT systems within the organisation.

# Purpose

The purpose of this policy is to define clear lines of accountability and responsibility for AI-driven decisions and actions within the organisation's IoT infrastructure. This policy aims to:

* Ensure that individuals and teams are held accountable for the development, deployment, and operation of AI systems.
* Promote transparency and explainability in AI-driven decision-making processes.
* Enable traceability of AI actions and outcomes back to responsible parties.
* Mitigate the risk of unintended consequences and ethical violations.

# Scope

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

# Policy Statement

## Clear Responsibility and Accountability

* **Designated Ownership:** Each AI model or system shall have a designated owner who is responsible for its development, deployment, operation, and maintenance.
* **Clear Roles and Responsibilities:** Roles and responsibilities for AI development, deployment, and oversight shall be clearly defined and documented.
* **Decision-Making Authority:** The authority to make decisions regarding the design, training, and deployment of AI models shall be clearly defined and assigned to appropriate personnel or teams.

## Documentation and Traceability

* **Model Documentation:** Comprehensive documentation shall be maintained for each AI model, including its purpose, design, training data, algorithms, and decision-making processes.
* **Audit Trails:** Audit trails shall be implemented to track and record AI-driven actions and decisions, enabling traceability and accountability.
* **Data Lineage:** The origin and lineage of data used for training and inference shall be documented to ensure transparency and accountability.

## Human Oversight and Control

* **Human-in-the-Loop:** Critical decisions or actions with significant impact shall be subject to human review and approval.
* **Override Mechanisms:** Mechanisms shall be in place to allow authorised personnel to override or intervene in AI systems if necessary to prevent harm or unethical outcomes.
* **Regular Reviews:** AI systems and their outputs shall be periodically reviewed to ensure their continued alignment with ethical principles and organisational objectives.

## Impact Assessments and Risk Management

* **Impact Assessments:** Prior to deployment, AI systems shall undergo impact assessments to evaluate their potential societal, ethical, and legal implications.
* **Risk Management:** Risks associated with AI systems, including the potential for bias, discrimination, or unintended consequences, shall be identified and mitigated through appropriate controls and safeguards.

## Ethical Considerations

* **Ethical Framework:** The organisation shall adhere to a defined ethical framework for the development and use of AI, ensuring that AI systems are aligned with the organisation's values and societal well-being.
* **Bias Mitigation:** Measures shall be taken to identify and mitigate potential biases in AI models and algorithms.
* **Transparency and Explainability:** AI systems shall be designed to provide transparent and understandable explanations for their decisions, especially in critical applications.

# Responsibilities

* **AI Ethics Committee:** An AI Ethics Committee shall be established to oversee the ethical development and use of AI within the organisation, including the promotion of accountability and responsibility.
* **Information Security Officer:** Responsible for ensuring that AI systems comply with this policy and relevant security standards.
* **Data Scientists and AI Developers:** Responsible for developing and deploying AI models in accordance with ethical principles and this policy.
* **Management:** Responsible for fostering a culture of ethical AI use and providing necessary resources and support.

# Breaches of Policy

Non-compliance with this policy may result in disciplinary action, up to and including termination of employment or contractual relationships. Additionally, the misuse or unethical use of AI systems may lead to reputational damage, legal consequences, and loss of trust among stakeholders.

# 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