TrustIoT Framework for Industry 4.0

"AI Ethics and Governance"

|  |  |
| --- | --- |
| 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](#_Toc176334429)

[2. Purpose 4](#_Toc176334430)

[3. Scope 4](#_Toc176334431)

[4. Policy Statement 4](#_Toc176334432)

[4.1. Ethical AI Principles 4](#_Toc176334433)

[4.2. Transparency and Explainability 4](#_Toc176334434)

[4.3. Accountability and Responsibility 4](#_Toc176334435)

[4.4. Fairness and Non-discrimination 5](#_Toc176334436)

[4.5. Human Oversight and Control 5](#_Toc176334437)

[5. Responsibilities 5](#_Toc176334438)

[6. Breaches of Policy 5](#_Toc176334439)

[7. Document Management 5](#_Toc176334440)

# Introduction

The integration of Artificial Intelligence (AI) into the Internet of Things (IoT) ecosystem presents opportunities for innovation and efficiency but also raises ethical concerns. The use of AI in decision-making processes, data analysis, and automation can have significant implications for individuals, society, and the organisation. This policy establishes a framework for ensuring the ethical and responsible development, deployment, and use of AI within the organisation's IoT infrastructure.

# Purpose

The purpose of this policy is to define the ethical principles and governance mechanisms that guide the organisation's use of AI in the IoT context. This policy aims to:

* Ensure that AI is used in a manner that aligns with the organisation's values and ethical principles.
* Promote transparency, fairness, and accountability in AI-driven decision-making.
* Mitigate the risk of bias, discrimination, and unintended consequences.
* Maintain human oversight and control over AI systems.

# Scope

This policy applies to all AI models, algorithms, and applications developed, deployed, or utilised within the organisation's IoT environment.

# Policy Statement

## Ethical AI Principles

* **Beneficence:** AI shall be used to promote human well-being, safety, and societal benefit.
* **Non-maleficence:** AI shall not be used to cause harm or inflict damage on individuals or society.
* **Autonomy:** AI shall respect human autonomy and avoid undue influence or manipulation.
* **Justice:** AI shall be fair and unbiased, avoiding discrimination or prejudice.
* **Explainability:** AI systems shall be transparent and provide understandable explanations for their decisions, especially in critical applications.

## Transparency and Explainability

* **Explainable AI (XAI):** The organisation shall strive to use explainable AI techniques to provide insights into the reasoning behind AI-driven decisions.
* **Documentation:** AI models and algorithms shall be documented, including their purpose, data sources, and decision-making processes.
* **Communication:** Clear and understandable explanations of AI-driven outcomes shall be provided to affected individuals and stakeholders.

## Accountability and Responsibility

* **Clear Responsibility:** Clear lines of responsibility and accountability shall be established for the development, deployment, and use of AI systems.
* **Human Oversight:** Human oversight and control mechanisms shall be implemented to ensure that AI systems operate within defined ethical boundaries.
* **Impact Assessments:** The potential impact of AI systems on individuals, society, and the environment shall be assessed before deployment.

## Fairness and Non-discrimination

* **Bias Mitigation:** Measures shall be taken to identify and mitigate potential biases in AI models and algorithms.
* **Fairness Testing:** AI systems shall be regularly tested to ensure fairness and avoid discriminatory outcomes.
* **Diverse Development Teams:** AI development teams shall be diverse and inclusive to promote a variety of perspectives and reduce the risk of bias.

## Human Oversight and Control

* **Human-in-the-Loop:** Critical decisions or actions involving AI shall be subject to human review and approval.
* **Override Mechanisms:** Mechanisms shall be in place to allow humans to override or intervene in AI systems if necessary.
* **Accountability:** Individuals and teams responsible for AI systems shall be held accountable for their actions and decisions.

# Responsibilities

* **AI Ethics Committee:** An AI Ethics Committee shall be established to oversee the ethical development and use of AI within the organisation.
* **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.

# 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