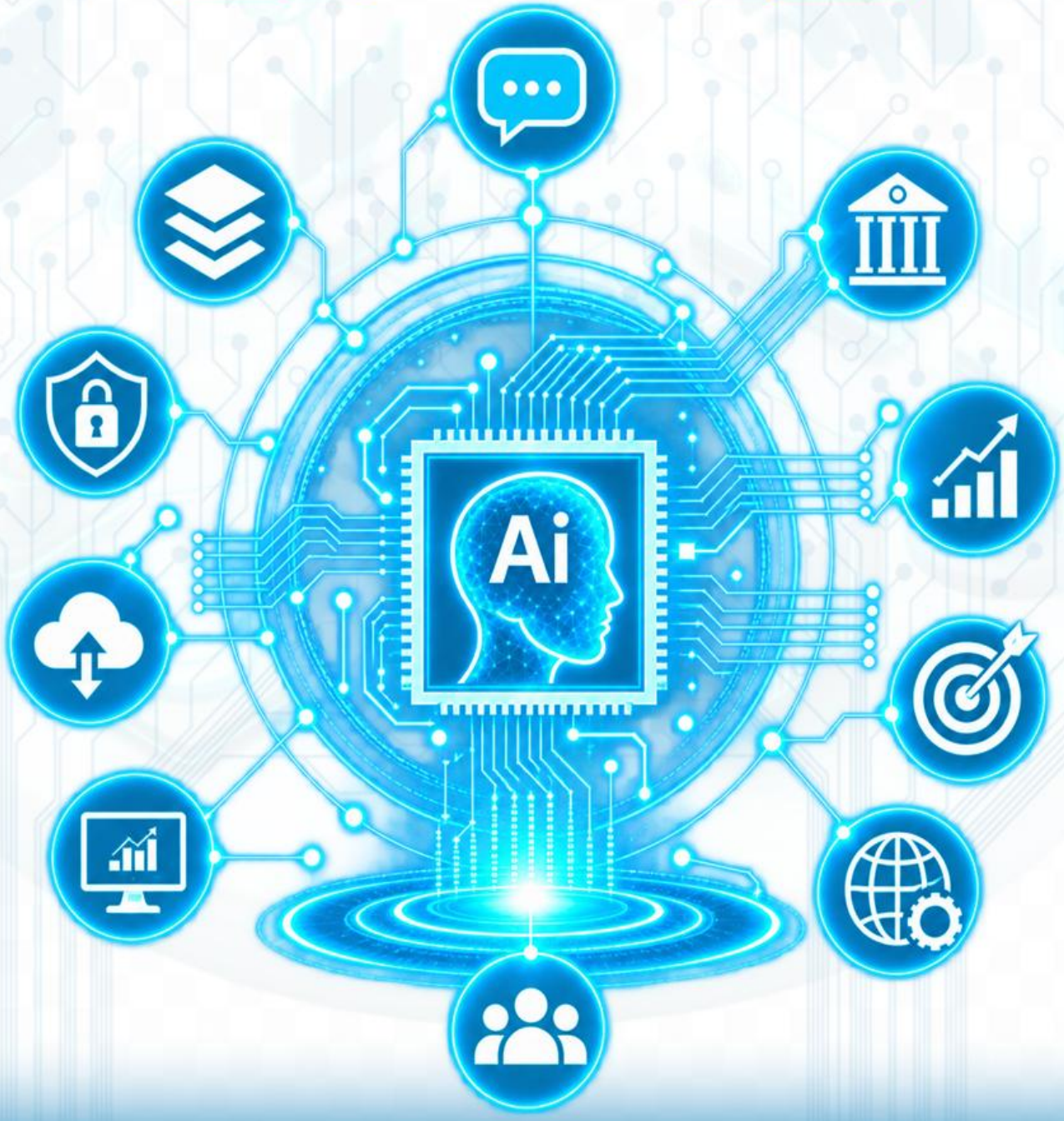
The Unted Republic of Tanzania

President's Office

e-Government Authority

ISO 9001:2015 | ISO/IEC 27001:2022 CERTIFIED

# Standards and Guidelines for Management, Implementation and Responsible Use of Artificial Intelligence (AI) in Public Institutions




August 2026

**THE UNITED REPUBLIC OF TANZANIA**

**PRESIDENT'S OFFICE**

**e-GOVERNMENT AUTHORITY**

**ISO 9001:2015|ISO/IEC 27001:2022 CERTIFIED**

**Document Title**

# Standards and Guidelines for Management, Implementation and Responsible Use of Artificial Intelligence (AI) in Public Institutions

**Document Number**

eGA/EXT/APA/011

| APPROVAL | Name | Job Title/ Role | Signature | Date |
|---|---|---|---|---|
| Approved by | Dr. Mussa M. Kissaka | Board Chairperson |  | 31.08.2026 |

## PREFACE

Artificial Intelligence (AI) is increasingly being adopted in Tanzania's public sector to improve service delivery, operational efficiency, and evidence-based decision-making. However, without appropriate governance, its use may introduce risks related to accountability, privacy, cybersecurity, bias, transparency, human oversight, and public trust. These risks require a coordinated governance framework to ensure the responsible, secure, and ethical adoption of AI across Public Institutions.

Pursuant to sections 5(2)(c), 5(2)(i), 5(2)(q), and 22(3) of the e-Government Act, Cap.273, R.E.2023, the Authority is mandated to provide guidance to Public Institutions on e-Government initiatives, set ICT standards and procedures, promote e-Government research, development and innovation. Furthermore, Dira 2050 embraces the use of emerging technologies such as Artificial Intelligence (AI) to accelerate digital transformation, enhance public sector efficiency, foster innovation, promote socio-economic progress, boost global competitiveness, and ensure ethical and accountable use, hence improving public service delivery.

In fulfilment of these requirements, the e-Government Authority has prepared these "*Standards and Guidelines for Management, Implementation and Responsible Use of Artificial Intelligence in Public Institutions"* to provide guidance on the responsible, secure, lawful, ethical, trustworthy, and accountable use of AI. These Standards and Guidelines operationalize the broader "e-Government Emerging Technologies Governance Framework".

The Authority urges all Public Institutions to adhere to these Standards and Guidelines when acquiring, developing or using AI systems or AI-enabled solutions. Proper application of this document will contribute to responsible AI use, strengthened public trust, protection of Government systems and data, and delivery of secure, inclusive, and accountable digital public services.

Dr. Mussa M. Kissaka
**BOARD CHAIRPERSON**

## Table of Contents

**PREFACE** ..... 3
**ACRONYMS** ..... 6
**GLOSSARY** ..... 7

**1 INTRODUCTION ..... 13**

**1.1 Overview** ..... 13
**1.2 Purpose** ..... 14
**1.3 Rationale** ..... 14
**1.4 Scope** ..... 15

**2 E-GOVERNMENT AI PRINCIPLES AND LIFE CYCLE ..... 16**

**2.1 e-Government AI Principles** ..... 16
**2.1.1 Public Value and Public Interest** ..... 16
**2.1.2 Fairness, Non-discrimination and Inclusion** ..... 16
**2.1.3 Accountability and Responsible Ownership** ..... 16
**2.1.4 Transparency, Explainability and Traceability** ..... 17
**2.1.5 Security, Safety and Resilience** ..... 17
**2.1.6 Privacy and Data Protection** ..... 17
**2.1.7 Interoperability and Integration** ..... 17
**2.1.8 Human Oversight and Control** ..... 17
**2.1.9 Sustainability** ..... 18
**2.2 e-Government AI Lifecycle** ..... 18

**3 STANDARDS AND GUIDELINES FOR MANAGEMENT, IMPLEMENTATION AND RESPONSIBLE USE OF ARTIFICIAL INTELLIGENCE (AI) IN PUBLIC INSTITUTION ..... 22**

**3.1 THE STANDARDS** ..... 22
**3.1.1 General Standards** ..... 22
**3.1.2 Specific Standards** ..... 23
**3.1.2.1 Initiation Standards** ..... 23
**3.1.2.2 Design Standards** ..... 23
**3.1.2.3 Acquisition Standards** ..... 29
**3.1.2.4 Development Standards** ..... 30
**3.1.2.5 Testing and Validation Standards** ..... 31
**3.1.2.6 Deployment Standards** ..... 35
**3.1.2.7 Operation and Monitoring Standards** ..... 35
**3.1.2.8 Retirement Standards** ..... 38
**3.2 THE GUIDELINES** ..... 39
**3.2.1 General Guidelines** ..... 39

**3.2.2 Specific Guidelines** ..... 41
**3.2.2.1 Initiation** ..... 41
**3.2.2.2 Design** ..... 42
**3.2.2.3 Acquisition Guidelines** ..... 43
**3.2.2.4 Development** ..... 47
**3.2.2.5 Testing and Validation Guidelines** ..... 48
**3.2.2.6 Deployment Guidelines** ..... 49
**3.2.2.7 Operation and Monitoring Guidelines** ..... 50
**3.2.2.8 Retirement Guidelines** ..... 52
**3.2.3 Guidelines on Prohibited AI Use Cases** ..... 53
**3.2.4 Guidelines for Responsible Use of Generative AI** ..... 54

**4 IMPLEMENTATION, REVIEW AND ENFORCEMENT ..... 56**
**5 RELATED DOCUMENTS ..... 56**
**6 DOCUMENT CONTROL ..... 57**
**APPENDIX I: POTENTIAL BENEFIT AND NEGATIVE IMPACTS ..... 59**
**APPENDIX II: IMPACT AREAS ..... 59**
**APPENDIX III: IMPACTS OF AI SYSTEM FAILURE ..... 64**
**APPENDIX IV: IMPACTS OF AI SYSTEM MISUSE OR ABUSE ..... 65**
**APPENDIX V: OVERALL AI SYSTEM IMPACT RATING ..... 65**

## ACRONYMS

| | |
|---|---|
| **AI** | Artificial Intelligence |
| **AIaaS** | Artificial Intelligence as a Service |
| **API** | Application Programming Interface |
| **AU** | African Union |
| **BLEU** | Bilingual Evaluation Understudy |
| **COTS** | Commercial Off-the-Shelf |
| **CPU** | Central Processing Unit |
| **e-GA** | e-Government Authority |
| **e-GSOC** | e-Government Security Operations Center |
| **GenAI** | Generative Artificial Intelligence |
| **GISP** | Government ICT Services Portal |
| **GPU** | Graphics Processing Unit |
| **HIC** | Human-in-Command |
| **HITL** | Human-in-the-Loop |
| **HOTL** | Human-on-the-Loop |
| **ICT** | Information and Communication Technology |
| **IEC** | International Electrotechnical Commission |
| **ISO** | International Organization for Standardization |
| **LLM** | Large Language Model |
| **NIST** | National Institute of Standards and Technology |
| **OECD** | Organisation for Economic Co-operation and Development |
| **PDPA** | Personal Data Protection Act |
| **PO-PSM** | President's Office - Public Service Management |
| **RBAC** | Role-Based Access Control |
| **ROUGE** | Recall-Oriented Understudy for Gisting Evaluation |
| **SLA** | Service Level Agreement |
| **UAT** | User Acceptance Testing |
| **UNESCO** | United Nations Educational, Scientific and Cultural Organization |

## GLOSSARY

| | |
|---|---|
| **Adversarial Testing** | The deliberate testing of an AI system using manipulated, malicious, or deceptive inputs designed to cause errors, in order to evaluate the system's robustness and security before and during operational use. |
| **Aggregation** | is the process of combining individual data points into summarized, group-level statistics so that insights can be derived from data without exposing any individual's specific information. |
| **AI as a Service** | The delivery of AI capabilities such as models, APIs, or platforms by an external provider over the internet on a subscription or usage basis, without the institution owning or hosting the underlying AI system. |
| **AI Impact Assessment** | A structured evaluation of how an AI system's deployment may affect individuals, groups, institutions, and society across dimensions such as fairness, privacy, safety, security, accountability, transparency, Explainability, and reliability. |
| **AI Incident** | An event in which an AI system's malfunction, misuse, compromise, or harmful output causes or has the potential to cause harm, service disruption, security exposure, or noncompliance, requiring detection, response, and escalation. |
| **AI solution** | A software-based solution that incorporates one or more AI systems or AI models to perform tasks that normally require human intelligence, such as prediction, classification, recommendation, decision support, perception, or content generation, in order to achieve specific business or operational objectives. |
| **AI System** | A machine-based system that, for explicit or implicit objectives, processes inputs to generate outputs such as predictions, content, recommendations, or decisions that can influence |

| | |
|---|---|
| | physical or virtual environments, operating with varying levels of autonomy. |
| **AI Use Case** | a specific, real-world application where AI solves a distinct problem. |
| **Anonymization** | is the process of permanently and irreversibly removing all identifying information from a dataset so that individuals can no longer be identified, directly or indirectly, under any circumstances. |
| **Artificial Intelligence** | is the capability of computer systems to perform tasks that normally require human intelligence, including learning, reasoning, prediction, content generation, decision support, and pattern recognition. |
| **Bias** | A systematic difference in the treatment or outcomes of certain individuals, groups, or objects compared to others, which may arise from training data, model design, or use context and may result in unfair or discriminatory outcomes. |
| **Black-Box AI** | An AI system whose internal workings, logic, or decision-making processes cannot be adequately inspected, understood, or explained by its users or operators. |
| **Commercial Off the Shelf (COTS) AI Solution** | ready-made product purchased from commercial vendors, typically licensed on per-user, per-seat, or enterprise basis. These solutions are pre-built, tested, and supported by the vendor, making them suitable for institutions with limited internal AI capacity. |
| **Concept Drift** | is the phenomenon where the statistical properties, patterns, or relationships in data change over time, causing an AI or machine learning model's performance, accuracy, or reliability to decline because the model was trained on historical data that no longer reflects current real-world conditions. |
| **Data Card** | A standardised document that provides key information about a dataset used in an AI system, including its sources, |

| | |
|---|---|
| | composition, collection methods, intended uses, limitations, and known gaps. |
| **Data Drift** | is the change in the statistical properties, distribution, structure, or characteristics of input data over time compared to the data originally used to train an AI or machine learning model. |
| **Data Leakage** | A situation in which information from outside the training dataset such as test data or future information improperly influences model training, producing misleadingly high-performance during evaluation that does not hold in real-world use. |
| **Edge case** | is a problem or situation that happens only under extreme, uncommon, or boundary conditions outside the normal range of everyday operations. |
| **Explainability** | The property of an AI system that allows the important factors influencing its outputs to be expressed in a way that humans can understand, including providing affected persons with sufficient information to understand and, where appropriate, challenge an outcome. |
| **Fairness Metric** | A quantitative measure of whether outputs or error rates differ unjustifiably between subgroups, such as demographic parity, disparate impact ratio, equal opportunity, or equalized odds. |
| **Feature Engineering** | The process of selecting, transforming, or creating input variables (features) from raw data to improve the performance and reliability of a machine learning model. |
| **Fine-tuning** | The process of further training a pretrained AI model on domain specific or task specific data so that it performs better in a particular context or environment. |
| **Generative AI (GenAI)** | is a transformative branch of AI that has the capability to generate new and original content, including images, code, text, speech, music, translations and more. |

| | |
|---|---|
| **Grounding** | Techniques used to anchor an AI system's outputs to verified, authoritative, or contextual data sources in order to reduce fabricated, harmful, or out-of-domain outputs. |
| **Hallucination** | The generation by an AI system, particularly generative AI, of content that is factually incorrect, fabricated, or unsupported by its input data or training, while being presented as accurate. |
| **Human-In-Command (HIC)** | is a governance mechanism that ensures humans maintain ultimate responsibility and control over AI systems throughout their lifecycle, including their deployment, operation, and outcomes. |
| **Human-in-the-Loop (HITL)** | A human oversight model in which a human must actively review and approve every AI model output before it is executed or acted upon. |
| **Human-on-the-Loop (HOTL)** | A human oversight model in which an AI system executes actions autonomously while a human monitor its operation in real time and retains the ability to intervene instantly. |
| **Kill Switch (Override Function)** | A secure, immediate control mechanism that allows authorized operators to pause or halt an AI system's operation and revert to manual control. |
| **Machine Learning** | A branch of artificial intelligence in which systems improve their performance on a task by learning patterns from data, rather than being explicitly programmed for every scenario. |
| **Masked data** | is a type of data in which sensitive information (such as Personally Identifiable Information (PII) or Protected Health Information (PHI)) has been modified, obscured, redacted, or replaced to protect privacy and security while preserving the usefulness of the data for AI model training, testing, analytics, development, or sharing. |
| **Model Card** | A standardized document accompanying an AI model that details its training parameters, intended use cases, |

| | |
|---|---|
| | performance benchmarks across demographic groups, and known limitations. |
| **Model Drift** | is the gradual decline in the performance, accuracy, reliability, or effectiveness of an AI or machine learning model over time due to changes in data patterns, operational environments, user behavior, or real-world conditions after the model has been deployed. |
| **Prohibited AI Use Case** | refers to any application or deployment of an artificial intelligence system that is banned or restricted because it creates unacceptable risks or threats to fundamental human rights, safety, privacy, fairness, democratic values, or societal well-being. |
| **Protected Attribute** | A personal characteristic such as race, gender, religion, age, or disability that must not be used as a basis for unfair or discriminatory treatment, and which requires special safeguards when processed by AI systems. |
| **Pseudonymization** | is the process of replacing identifying information with artificial identifiers called pseudonyms, so that data can no longer be attributed to a specific individual without the use of additional separately stored information. |
| **Residual Risk** | The level of risk that remains after mitigation measures, controls, or safeguards have been applied. |
| **Robustness** | The ability of an AI system to maintain safe, reliable, and accurate performance under variable, unexpected, adverse, or adversarial conditions, including abnormal inputs and system disruptions. |
| **Synthetic data** | is artificially generated data that mimics the statistical properties and patterns of real-world data, without containing any actual real records or personal information. |
| **Tokenization in Privacy** | is the process of replacing sensitive data elements with a randomly generated, non-sensitive placeholder called a token, |

| | |
|---|---|
| | which has no exploitable meaning or value outside the tokenization system. |
| **Training, Validation and Test Data** | The three separated datasets used in AI development: training data is used to teach the model, validation data is used to tune and compare models during development, and test data is kept fully isolated to provide an independent, unbiased evaluation of final model performance. |
| **Use Case** | a practical, real-world scenario in which a product, service, or system is used by a person (or another system) to achieve a specific goal. |
| **Vendor Lock In** | A situation in which an institution becomes dependent on a specific vendor's AI products, services, or platforms and cannot switch to alternatives without substantial cost, disruption, or loss of capability. |

# 1 INTRODUCTION

## 1.1 Overview

The e-Government Authority (e-GA) is an ISO 9001: 2015 and ISO/IEC 27001:2022 certified Public Institution established under e-Government Act, Cap.273, R.E.2023 mandated to coordinate, oversee, and promote e-Government initiatives as well as enforce e-Government related policies, laws, regulations, standards, and guidelines in public institutions.

The increasing adoption of Artificial Intelligence (AI) presents significant opportunities to enhance public service delivery, operational efficiency, innovation, and evidence-based decision-making. However, AI systems also introduce specific risks associated with data dependency, model opacity, probabilistic outputs, bias, data and model drift, vendor dependency, unsafe automation, cybersecurity, privacy, and overreliance on AI-generated or AI-assisted outputs. Without appropriate governance, these risks may affect the security, reliability, fairness, accountability, and trustworthiness of AI-enabled public services. The document therefore establishes general and lifecycle-specific standards and guidelines covering AI initiation, design, acquisition, development, testing and validation, deployment, operation and monitoring, and retirement.

These Standards and Guidelines for Management, Implementation and Responsible Use of Artificial Intelligence (AI) in Public Institutions operationalize the broader e-Government Emerging Technologies Governance Framework by translating its governance principles into AI-specific standards, requirements, and practical guidance. They provide a structured basis for managing AI systems and AI-enabled solutions throughout their lifecycle. In addition to the lifecycle requirements, the Standards and Guidelines provide specific provisions for AI risk and impact management, prohibited AI use cases, and the responsible use of Generative AI.

The Standards and Guidelines are therefore intended to promote responsible innovation and sustainable adoption of AI in the public sector, while strengthening institutional governance, public trust, protection of Government systems and data, and the delivery of secure, inclusive, efficient, reliable, and accountable digital public services.

## 1.2 Purpose

The purpose of this document is to provide guidance in management and implementation of Artificial Intelligence (AI) to ensure ethical, secure, trustworthy and responsible use of AI in public institutions.

## 1.3 Rationale

Artificial Intelligence (AI) have been adopted and used by public institutions to support public service delivery through capabilities such as forecasting, classification, recommendation, automation, natural language processing, image analysis, anomaly detection, and decision support. However, the adoption and use of AI without proper governance and management may result in increased security, privacy, ethical, and operational concerns.

In line with sections 5(2)(c), 5(2)(i), 5(2)(q), and 22(3) of the e-Government Act, Cap.273, R.E.2023, the Authority is mandated to provide guidance to Public Institutions on e-Government initiatives, set ICT standards and procedures, promote e-Government research, development and innovation. Moreover, the use of emerging technologies, including Artificial Intelligence (AI), is recognized under Dira 2050 as an important enabler of digital transformation, public sector efficiency, innovation, socio-economic development, global competitiveness, and improved public service delivery through ethical and accountable technology use.

Artificial Intelligence requires specific standards and guidelines to harness its benefits while effectively managing risks arising from areas such as data dependency, model opacity, automated recommendations, probabilistic outputs, generative capabilities, bias, drift, vendor dependency, unsafe automation, and overreliance on AI outputs. In the absence of AI-specific standards and guidelines, AI systems may be adopted inconsistently, resulting in lack of ownership, inadequate documentation, unclear accountability, insufficient testing, poor data governance, weak human oversight, cybersecurity exposure, unreliable outputs, uncontrolled use of generative AI, and reduced public trust.

These Standards and Guidelines therefore provide practical requirements and guidance to support responsible, secure, lawful, ethical, trustworthy, inclusive, interoperable, and accountable AI implementation in Public Institutions.

## 1.4 Scope

This document applies to all public institutions and covers all AI systems, AI enabled solutions and AI capabilities embedded within other systems that are developed, acquired, deployed, managed, or used by the public institution in support of its operations. It also applies to existing Government systems that are currently enhanced with AI capabilities. Moreover, the standards and guidelines apply to all relevant processes throughout the AI lifecycle.

# 2 e-GOVERNMENT AI PRINCIPLES AND LIFE CYCLE

Artificial Intelligence (AI) is the capability of computer systems to perform tasks that normally require human intelligence, including learning, reasoning, prediction, content generation, decision support, and pattern recognition. AI systems operate with different levels of autonomy, ranging from simple rule-based systems to advanced models that can learn and improve over time. Depending on the design and use, AI systems can support or automate decision-making processes across various public sectors such as healthcare, agriculture, education, and mining.

## 2.1 e-Government AI Principles

The e-Government AI principles set out in this section are derived from the governance principles established under the e-Government Emerging Technology Framework and adapted to the specific context of Artificial Intelligence use in Public Institutions.

### 2.1.1 Public Value and Public Interest

This principle emphasizes that AI systems should be adopted to support legitimate public functions, institutional needs, service delivery objectives, or public-interest outcomes. It ensures that AI adoption is driven by clear public value, necessity, and proportionality rather than novelty, market pressure, automation bias, or vendor influence.

### 2.1.2 Fairness, Non-discrimination and Inclusion

This principle emphasizes that AI systems are designed, acquired, developed, deployed, and used in a manner that avoids unfair treatment, discriminatory outcomes, unjust exclusion, arbitrary decisions, or disproportionate adverse effects on individuals, groups, communities, or users of public services. It ensures that AI-supported public service delivery remains lawful, ethical, inclusive, and consistent with institutional mandates.

### 2.1.3 Accountability and Responsible Ownership

This principle emphasizes that responsibilities for each AI system are clearly assigned throughout AI lifecycle. It recognizes that AI models, automated outputs, vendors, developers, technical teams, or system users do not replace institutional accountability for decisions, actions, outcomes, risks, and consequences arising from the use of AI in public-service delivery.

### 2.1.4 Transparency, Explainability and Traceability

This principle emphasizes that AI systems should operate in a manner that enables appropriate understanding of their purpose, functionality, data use, outputs, limitations, and decision-support role. It ensures that AI-supported processes are understandable to relevant stakeholders and that key actions, data flows, outputs, approvals, and system changes can be traced and audited throughout the AI lifecycle.

### 2.1.5 Security, Safety and Resilience

This principle emphasizes the protection of Government systems, data, services, users, and operations from compromise, misuse, malfunction, harmful AI outputs, service disruption, and continuity failure. It supports secure, safe, risk-based, and resilient AI adoption through appropriate safeguards, monitoring, incident response, recovery, and lifecycle controls.

### 2.1.6 Privacy and Data Protection

This principle emphasizes the lawful, secure, purpose-bound, and accountable handling of data and information processed by AI systems. It is particularly important where AI systems use personal, sensitive, confidential, or institutional data to support public-service delivery, analysis, automation, prediction, recommendation, or decision support.

### 2.1.7 Interoperability and Integration

This principle emphasizes that AI systems are designed, acquired, developed, deployed, and operated to work securely and effectively with Government systems, data platforms, devices, workflows, and service delivery channels. It supports secure data exchange, standardization, compatibility, maintainability, and coordinated service delivery, while reducing duplication, fragmentation, and isolated AI deployments.

### 2.1.8 Human Oversight and Control

This principle emphasizes the involvement of qualified individuals in supervising, monitoring, and managing AI systems and their outputs. It ensures that AI-supported decisions, actions, and processes remain lawful, ethical, accurate, explainable, and aligned with institutional mandates, public-service objectives, and approved use conditions.

### 2.1.9 Sustainability

This principle emphasizes that AI systems are developed, acquired, deployed, and used in a manner that supports innovation, efficiency, continuous improvement, and long-term public value, while minimizing negative social, economic, environmental, and institutional impacts. It ensures that AI adoption remains resource-efficient, resilient, maintainable, and aligned with sustainable national development goals. It also emphasizes continuous capacity building and professional development to equip staff with the knowledge, skills, and competencies necessary to effectively manage, use, maintain, and govern AI solutions throughout their lifecycle.

## 2.2 e-Government AI Lifecycle

The e-Government AI lifecycle provides step-by-step guidance on how public institutions should plan, build, acquire, test, deploy, operate, and eventually retire AI systems. It ensures that AI is introduced into public services in a responsible, well-governed, and accountable way.

The lifecycle consists of seven phases, each phase has a clear purpose, a set of required activities, and governance requirements that must be met before moving to the next stage. While the phases are presented in order, the lifecycle is not strictly one-directional, institutions may need to revisit earlier phases as new risks, requirements, or material changes emerge. AI Impact Assessment, risk management and AI principles apply throughout all phases. Figure 1 illustrates the seven phases of the e-Government AI Lifecycle and how they connect to form a continuous cycle. Table 1 provides a summary of each phase.

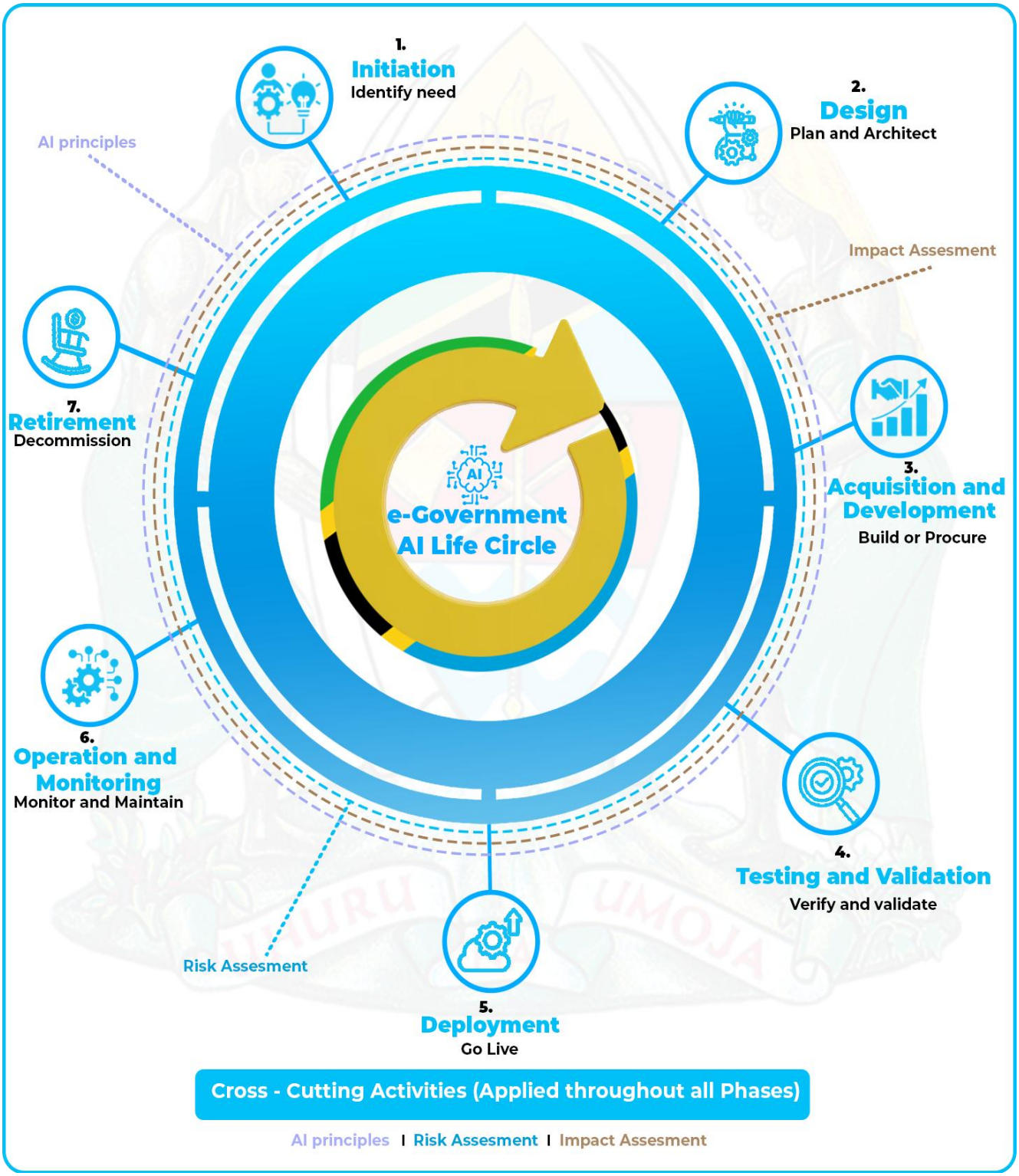



Figure 1: e-Government AI Lifecycle

Table 1: Summary of e-Government AI Lifecycle

| S/N | Phase | What Happens |
|---|---|---|
| 1 | Initiation | The institution identifies a problem or service need that AI could address. The AI use case is defined, covering the problem, intended benefit, proposed solution, affected persons, data categories, and AI's role. An AI Impact Assessment and AI Risk Assessment are conducted. Both the use case and assessment reports are submitted to the ICT Steering Committee for approval, then to e-GA through the Government ICT Services Portal for clearance. |
| 2 | Design | The approved AI use case is translated into a detailed technical plan covering system architecture, data design, model selection, explainability, fairness, security, human oversight, integration, sustainability, and testing. AI systems must support meaningful human control (HITL, HOTL or HIC models) and include a kill-switch or override function. Data design must specify sources, legal basis, quality requirements, and protection measures. |
| 3 | Acquisition & Development | The AI system is obtained through in-house development, open-source adoption, commercial procurement, outsourced development, or AI as a Service. Public funds must be protected, citizen data safeguarded, and the system must remain under government control. For procurement, contracts must define ownership, audit rights, and knowledge transfer. e-GA approval is required before acquiring AI as a Service. |
| 4 | Testing & Validation | The AI system is thoroughly tested, evaluated, verified, and validated before going live. Testing covers functional performance, robustness, bias, fairness, privacy, cybersecurity, adversarial scenarios, explainability, human oversight, integration, and user acceptance. Test data must be representative, separated from training data, and assessed for bias. All results are documented in a Testing and Validation Report required before deployment authorization. |

| | | |
|---|---|---|
| 5 | Deployment | The AI system moves from testing into the production environment. A deployment readiness assessment and deployment plan are prepared, covering scope, safeguards, success criteria, monitoring, and exit conditions. The hosting environment is assessed and confirmed as production-ready. The system is formally handed over to the operational team. Where AI affects public services or citizen rights, fallback procedures and appeal mechanisms must be in place from day one. |
| 6 | Operation & Monitoring | The longest phase of the lifecycle. The AI system is continuously monitored for availability, performance, security, safety, and reliability. Mechanisms detect anomalies, model drift, bias, harmful behavior, and security threats. A feedback mechanism is in place for users to report issues. AI incidents are managed through institutional procedures with e-GA escalation. |
| 7 | Retirement | The planned and orderly shutdown of an AI system, triggered when it no longer meets requirements, presents unacceptable risks, becomes obsolete, or is replaced. A detailed retirement plan is developed covering scope, timelines, data handling, access revocation, infrastructure decommissioning, and record retention. Retirement must follow change management procedures and be approved by the ICT Steering Committee. |

# 3 STANDARDS AND GUIDELINES FOR MANAGEMENT, IMPLEMENTATION AND RESPONSIBLE USE OF ARTIFICIAL INTELLIGENCE (AI) IN PUBLIC INSTITUTION

## 3.1 THE STANDARDS

### 3.1.1 General Standards

Public Institution intending to adopt, develop or use AI shall ensure that:

i. the institutional ICT Steering Committee provides the overall AI Governance within the institution;

ii. AI is applied only where its use is justified, proportionate and demonstrably beneficial compared to non-AI alternatives;

iii. comprehensive documentations are maintained throughout the AI lifecycle including but not limited to:

   a) design and system architecture specification;
   b) design choices made and quality measures taken during the system development process;
   c) information about the data used during system development;
   d) assumptions made and quality measures taken on data quality such as assumed statistical distributions;
   e) management activities such as risk management taken during development or operation of the AI system;
   f) testing and validation records;
   g) changes made to the AI system when it is in operation; and
   h) impact assessment documentation.

iv. AI Impact Assessment and AI Risk Assessment are conducted and maintained throughout the AI lifecycle, including the identification of risk scenarios, impacts, implementation of mitigation measures, and continuous monitoring of residual risks;

v. affected persons are informed where AI is used to interact with them, prioritize them, recommend actions about them, assess eligibility, support decisions, or generate

official communication, where required by law, policy, risk classification, or public service context;

vi. visual watermarks and/or associated metadata are embedded in AI-generated media to ensure transparency and verifiable origin, including clear attribution of authorship; and

vii. AI systems operate in accordance with the principle of least privilege by granting only the minimum access rights, permissions, and privileges necessary to perform their approved functions.

### 3.1.2 Specific Standards

#### 3.1.2.1 Initiation Standards

Public Institution shall:

i. identify and document AI use cases with their business and operational justification;

ii. perform Impact Assessment and Risk Assessment of the AI use case proportionate to the nature, purpose, and context of the use case before approval for development, acquisition or operational use; and

iii. secure approval from institutional ICT Steering Committee and clearance from the e-Government Authority prior to the implementation of the AI use case.

#### 3.1.2.2 Design Standards

**a. Generic Design Standards**

Public institution shall design AI systems that:

i. are inclusive and meet accessibility requirements;

ii. respect human rights, promote fairness and align with applicable legal, ethical and societal values;

iii. support effective human interaction, clarity of outputs, interpretability and ease of use;

iv. support meaningful human oversight and control, including the ability for authorized personnel to intervene, override or suspend system operation where appropriate;

v. operate reliably and accurately under real world conditions considering variability in data, user behavior and environmental factors;

vi. prevent, detect and mitigate reasonably foreseeable physical and non-physical harms throughout the AI system lifecycle; and

vii. protect the confidentiality, integrity, availability and privacy of information throughout the AI system lifecycle.

**b. Data Design and Governance**

Public Institution shall:

i. define and document the data sources, categories of data, quantity of data, data subject, data rights, associated metadata and provenance of data used by the AI system;

ii. identify data that are relevant, lawful, proportionate and fit for AI system's intended purpose;

iii. adhere to privacy and data protection principles including data minimization, purpose limitation and appropriate security controls;

iv. define and adheres to data quality requirements for completeness, accuracy, consistency, timelines, validity and representativeness;

v. identify and address class imbalance, sampling bias, and disproportionate representation within datasets, where applicable, to reduce the risk of biased, unreliable, or systematically inaccurate AI outcomes;

vi. include datasets that are sufficient in volume, appropriate in variety, timely where required, reliable, and suitable for the AI system's intended purpose and deployment context, proportionate with the AI system's risk profile;

vii. define dataset requirements to ensure adequate representation of normal operating conditions, boundary conditions, edge cases, and reasonably foreseeable rare scenarios relevant to the AI system's intended purpose and operational context; and

viii. define appropriate partitioning of training, validation, and test datasets to support independent model training, validation, and performance evaluation.

### c. Model Selection and Architecture

Public institution shall ensure that:

i. AI models are selected based on, but not limited to the following considerations:
   a. appropriateness of the model for the use case;
   b. the complexity of the model and number of parameters;
   c. the ability to align, interpret and explain the model outputs;
   d. characteristics of training datasets including size, integrity, quality, sensitivity, age, relevance and diversity; and
   e. performance and auditability.

ii. the model architecture is appropriately documented including components, dependencies, data flows, interfaces and control points.

### d. Explainability and Transparency by Design

Public Institution shall ensure that the AI system design:

i. incorporates a mechanism for users to be informed when they are interacting with AI system or receiving AI assisted output;

ii. provides mechanisms to explain AI outputs to the intended users at a level of detail appropriate to their roles, responsibilities, and the intended use of the AI system;

iii. incorporates explainability capabilities that provide appropriate information to authorized technical personnel to enable understanding of the AI system's behavior, outputs and decision-making processes, commensurate with the AI system's intended purpose and risk profile, to support system development, testing, verification, validation, monitoring, maintenance, auditing, and incident investigation;

iv. facilitates explainability of any automated recommendation or prediction that could influence a consequential decision; and

v. incorporates mechanisms to record inference events and relevant metadata, commensurate with the AI system's intended purpose and risk profile, to support traceability, reproducibility, auditing, monitoring, and incident investigation.

### e. Fairness and Bias Mitigation

Public Institution shall ensure that:

i. the dataset reflects the demographic diversity of the intended population, where human-related data are used, to reduce the risk of biased or disproportionately poor performance across population groups;

ii. fairness objectives, appropriate fairness evaluation metrics, and acceptance criteria are defined and documented before model development or selection proportionate to the AI system's intended purpose and risk profile;

iii. protected attributes such as race, gender, religion and age are not used as direct model inputs unless their use is authorized by applicable law or is demonstrably necessary for the AI system's intended purpose and supported by documented justification and appropriate safeguards;

iv. subgroup or protected attribute labels required for fairness evaluation are maintained within validation and test datasets, even when those attributes are excluded from model inputs, in accordance with applicable privacy, data protection, and information security requirements; and

v. where fairness assessments identify unacceptable bias or disparities, the AI system design shall incorporate appropriate measures to mitigate the identified risks before deployment.

### f. Security and Robustness

Public Institution shall:

i. incorporate security-by-design and secure-by-default principles into the AI system design to protect the confidentiality, integrity, availability and authenticity of the AI system, associated information assets, and supporting infrastructure throughout the AI system lifecycle; and

ii. design AI systems to ensure robustness, resilience, and fault tolerance, with appropriate continuity measures to maintain safe, reliable, and secure functionality under unexpected conditions, failures, or system disruption.

### g. Human Oversight and Control

Public Institution shall ensure that:

i. the system design document specifies the human oversight model appropriate to the AI system' intended purpose and risk profile including where applicable:
   a. Human-in-the-Loop (HITL) where a human must actively review and approve every model output before specified actions are taken;
   b. Human-on-the-Loop (HOTL) where the system executes actions autonomously but a human monitor operation and can intervene instantly by modifying, suspending, or restricting AI solution where necessary; or
   c. Human-in-Command (HIC) where a human retains ultimate authority and accountability for the AI system by overseeing its objectives, deployment, operation, governance, and use, including the authority to modify, suspend, or terminate its operation.

ii. escalation procedures and decision-making responsibilities are defined for established risk thresholds, anomalous conditions, and situations requiring human intervention;

iii. the system design includes a secure override function that allows authorized operators to pause the AI model; and

iv. the system design identifies clear roles, responsibilities, and accountability for human oversight activities.

### h. System Integration and Interoperability

Public Institution shall ensure that:

i. AI systems are designed to integrate securely and interoperable with relevant information systems, data platforms, services, and business processes, consistent with the AI system's intended purpose;

ii. system integration requirements, data schemas and system interfaces including API, database connections, messaging services and file exchange mechanisms are documented, version controlled, and maintained throughout the AI system lifecycle; and

iii. integration points preserve the confidentiality, integrity, availability, and authenticity of information by implementing security controls appropriate to the AI system's intended purpose, risk profile, and operational environment.

### i. Sustainability Consideration

Public Institution shall ensure that the AI System Design includes:

i. requirements for scalability, maintainability, computational efficiency, resource utilization (compute, storage, power, cooling) and environmental sustainability; and

ii. mechanisms to detect and monitor data drift, concept drift, model drift, and other changes that could adversely affect the AI system's performance, reliability, or intended operation.

### j. Testing and Validation Planning

Public institution shall:

i. prepare an AI testing and validation plan, supported by detailed test cases and procedures, that, at a minimum, defines the testing scope and objectives, measurable acceptance criteria, appropriate and representative test data, testing under normal and adverse conditions, and assessments of fairness and bias, security and privacy, explainability, and human oversight;

ii. define how probabilistic, non-deterministic or context-dependent outputs will be evaluated against expected outcomes, acceptable tolerances, repeated evaluations, confidence thresholds and appropriate statistical or human-evaluation methods;

iii. design the AI system to support component, integration and end-to-end testing, including the testing of data pipelines, interfaces and, where applicable, prompts, thresholds and decision rules; and

iv. define the material changes or conditions requiring re-testing or revalidation, including model updates, retraining, changes to data sources, prompts, thresholds

or integrations, data or concept drift, performance degradation, security incidents and changes to the intended purpose or operating environment.

**k. Documentations**

Public institution shall prepare, maintain and keep up to date all documentation required under the Standards and Guidelines for Development, Acquisition, Operation and Maintenance of e-Government Application. In addition, the institution shall maintain the following AI-specific documentation where applicable:

i. documentation for each AI model, such as a Model Card that describe the model's intended purpose, training methodology, training parameters, performance characteristics, evaluation results, known limitations, assumptions, and intended operating conditions;

ii. documentation for each dataset, such as a Data Card that describe its source, purpose, composition, quality characteristics, collection method, preprocessing activities, licensing conditions, usage restrictions, and known limitations; and

iii. documentation justifying the selection of the AI architecture, algorithm, model, configuration settings, hyperparameters, and tuning approach, where applicable, based on the approved use case, data characteristics, acceptance criteria, and assessed risks.

#### 3.1.2.3 Acquisition Standards

Public institution shall:

a. ensure adherence to Public Procurement Act and its Regulations during the acquisition of AI systems;

b. ensure that acquisition and contracting documentation addresses:

i. the intended purpose and scope of the AI use case;

ii. the roles and responsibilities of the provider and the institution;

iii. privacy, security, and data handling requirements;

iv. documentation, transparency, traceability, and cooperation requirements;

v. service levels agreements (SLA), support arrangements, and incident notification requirements;

vi. hosting, access, continuity, and handover requirements;

vii. interoperability and integration obligations;

viii. change notification and approval requirements; and

ix. suspension and termination arrangements.

c. conduct AI acquisitions in a way that is problem-driven, transparent, risk-aware, and fair, ensuring that public benefit is prioritized and that the process supports long-term accountability and knowledge transfer;
d. acquire AI systems in a way that protects public funds, promotes fair competition, safeguards citizen data, and ensures that AI systems continue to perform well and remain under government control throughout their entire lifecycle;
e. manage AI acquisition projects from problem definition through to long-term sustainment, ensuring clear requirements, proper data management, appropriate contract arrangements, rigorous testing, ethical compliance, and correct funding allocation at every stage;
f. conduct due diligence before acquiring, subscribing to, relying on, or materially integrating a third-party AI model, platform, system, service, or provider;
g. ensure that the acquired AI system undergoes security assessment before implementation or operational use, including the assessment of the system's security controls, data protection measures, privacy safeguards, supply chain risks, AI-specific security risks, and compliance with applicable legal, regulatory, and organizational requirements; and
h. assess the risk of vendor lock-in ensuring that procurement decisions consider interoperability, portability, data ownership, and exit strategies appropriate to the institution's operational and business continuity requirements.

#### 3.1.2.4 Development Standards

Public institution shall:

a. establish and implement comprehensive data governance to ensure the secure, reliable, and well-coordinated flow of data, while maintaining its integrity, traceability, and proper handling throughout its entire lifecycle;
b. implement and document standardized processes for data preprocessing including data cleaning, selection, transformation, feature engineering, feature selection, validation, and integration to ensure that data is accurate, consistent, reliable, and fit for its intended purpose;

c. establish and manage AI model datasets that are representative, unbiased, and properly segregated, while incorporating contextual data to ensure accurate, reliable, and relevant AI outputs;
d. train models in a secure hosting environment with adequate capacity and resources to support the required AI operations;
e. implement systematic model creation, tuning, and grounding processes, ensuring that models are optimized for their target domains and governed by strict performance metrics and output safety boundaries;
f. establish a formal process for evaluating and comparing multiple trained AI models to ensure that the most suitable model is selected for deployment based on defined performance criteria;
g. develop AI systems with robust security measures; and
h. ensure comprehensive model and source code management as stipulated in Section 2.2.2.4 of the Standards and Guidelines for Development, Acquisition, Operation and Maintenance of e-Government Applications.

#### 3.1.2.5 Testing and Validation Standards

**a. Testing Plan**

Public Institution shall:

i. prepare a testing plan before testing, piloting, deploying, or materially modifying an AI system;
ii. ensure that the testing plan defines the approved AI use case, intended purpose, test objectives, test scope, test methods, success criteria, responsible roles, approval checkpoints, and required evidence;
iii. identify the functional, non-functional, legal, security, privacy, operational, human oversight, and accountability requirements against which the AI system shall be tested, evaluated, verified, and validated;
iv. define performance metrics and acceptable thresholds applicable to the AI system, including accuracy, error rates, precision, recall, confidence levels, throughput, BLEU/ROUGE Scores and any other metrics relevant to the AI use case;

v. ensure that test plan includes the testing domains applicable to the AI system such as functional testing, performance testing, robustness testing, bias and fairness testing, privacy testing, cybersecurity testing, adversarial testing, explainability testing, human oversight testing, integration testing, and user acceptance testing;

vi. ensure that the testing plan covers the complete AI-enabled service, including the AI model, supporting data, system integrations, business processes, human oversight mechanisms, security controls, and the intended operational environment, rather than evaluating the AI model or algorithm in isolation; and

vii. document test data, test scenarios, test scripts, test conditions, test tools, test environments, test limitations, and responsible officers.

### b. Test Data and Test Environment

Public Institution shall ensure that:

i. test data description covering data source, collection method, coverage, quality, representativeness, sensitivity, limitations, and known gaps are prepared;

ii. test and validation data are relevant, representative, sufficient, accurate, and suitable for the intended purpose and deployment context of the approved AI use case;

iii. test data are assessed for bias, imbalance, missing values, outdated records, outliers, inappropriate proxies, data leakage, quality defects, and mismatch with the expected production environment;

iv. test data are protected throughout their lifecycle by implementing appropriate administrative, technical, and physical controls based on the data's classification, sensitivity, and associated risks;

v. test data are separated from training data and validation data to ensure independent, unbiased, and reliable evaluation of AI model performance, accuracy, and capability;

vi. personal data used in testing are processed lawfully, fairly, securely, and only for explicit, specified, and legitimate purposes;

vii. anonymized, pseudonymized, synthetic, masked, or minimized data are used for testing where such data are sufficient to meet the testing objective and reduce privacy, confidentiality, or security risk; and

viii. test environments are isolated, secured, monitored, and controlled to prevent unauthorized access, leakage, misuse, contamination, or interference with production systems.

### c. System Testing

Public Institution shall:

i. test whether the AI system performs as per approved requirements, specifications, controls and operational procedures;
ii. test whether the AI system integrate correctly based on integration and interoperability requirements;
iii. test and evaluate the AI system against defined performance metrics and acceptance thresholds;
iv. test for robustness, material performance variation, known limitations, failure modes, edge cases, abnormal inputs, adversarial examples, and foreseeable misuse scenarios;
v. test whether mechanisms for detecting and reporting data drift, model drift and concept drift are effective;
vi. test AI systems to identify bias, discriminatory effects, unjustified exclusion, and material output variations that may result in unfair treatment, inappropriate differential impact, or unjustified adverse outcomes for affected persons or groups;
vii. test whether the AI system complies with applicable privacy, personal data protection, confidentiality, data minimization, retention, access control, and data subject rights requirements before operational use;
viii. test the AI system against relevant security and adversarial risks that may affect confidentiality, integrity, availability, resilience, or safe operation;
ix. test whether approved human oversight arrangements operate effectively, including the ability of authorized officers to review, interpret, override, reject, escalate, suspend, or disregard AI outputs where necessary; and
x. test explainability by reviewing sample AI outputs to determine whether the reasons, key influencing factors, limitations, confidence level where applicable, and required

human review actions are clear, understandable, and appropriate for the intended users and risk level of the AI system.

### d. Validation

Public Institution shall:

i. assess that test and validation data are lawful, relevant, representative, accurate, complete, timely, traceable, properly separated, and fit for the approved AI use case;
ii. assess test and validation data for bias, imbalance, missing values, outdated records, inappropriate proxies, data leakage, and mismatch with the intended deployment context;
iii. document data validation results, including data limitations, coverage gaps, quality issues, and any required corrective action;
iv. validate that the model, model version, architecture, assumptions, limitations, and intended use are documented and suitable for the approved AI use case;
v. evaluate the model against approved performance metrics and acceptance thresholds, including accuracy, precision, recall, false positive rate, false negative rate, latency, availability, or other metrics relevant to the AI use case;
vi. assess whether mechanisms for detecting, monitoring, reporting, escalating, and responding to drifts including data, model, and concept are established and operating where applicable;
vii. document model validation results, including model suitability, limitations, residual risks, corrective actions, and re-testing triggers; and
viii. ensure that validation activities are conducted or reviewed with sufficient independence from development, configuration, procurement, or supplier activities, corresponding with the risk and potential impact of the AI use case.

### e. Testing and Validation Documentation

Public Institution shall:

i. prepare and retain a Testing and Validation Report and supporting evidence before deployment authorization is requested;

ii. ensure that the Testing and Validation Report records tested system version, the testing performed, responsible officers, key results, recommended next step and re-testing triggers; and

iii. ensure that unresolved Testing and Validation findings are tracked to closure before deployment or formally accepted as residual risk by the Institutional ICT Steering Committee.

#### 3.1.2.6 Deployment Standards

Public Institution shall:

i. conduct a deployment readiness assessment before AI system is placed into production environment;

ii. define a deployment plan to include the scope, duration, users, safeguards, success criteria, monitoring requirements, restrictions and exit conditions for each pilot or phased deployment;

iii. obtain deployment authorization from the Institutional ICT Steering Committee before deploying an AI system into production; and

iv. ensure that deployment is phased, piloted or introduced in a controlled manner proportionate to the AI use case.

#### 3.1.2.7 Operation and Monitoring Standards

**a. AI system Availability and Performance**

Public Institution shall:

i. monitor AI system availability, performance, safety, integrity, reliability and capacity against defined service levels;

ii. implement performance baselines and thresholds to detect and address degradation proactively; and

iii. conduct regular capacity review and plan of the AI infrastructure to support current and future demand.

**b. AI Incident Management**

Public Institution shall:

i. Detect, log, classify, and resolve AI system incidents in a timely manner;
ii. Perform root cause analysis for recurring or high-impact AI incidents; and
iii. Maintain a knowledge base of AI system incidents to support faster incident resolution.

### c. AI System maintenance

Public institution shall ensure that:

i. changes to the AI System adhere to the Institutional Change Management Procedures, clearly stating the nature of the change, justification, affected AI components (such as models, datasets, or pipelines), impact analysis and requester information;
ii. changes to AI system are initiated only for valid reasons such as model performance improvement, bias correction, data drift response, security fixes, infrastructure enhancement, and operational issues detected through monitoring or user feedback;
iii. prioritizes approved requests based on urgency, risk level, impact on model performance, and potential ethical or operational implications before proceeding to impact analysis;
iv. the AI maintenance team conducts a detailed analysis of each accepted change request to assess the feasibility and potential impact on the AI system, including model behavior, training data, user experience, decision-making processes, and related applications, as well as risks such as bias amplification, performance degradation, data quality issues, security vulnerabilities, and compliance with ethical and regulatory requirements;
v. the required resources are estimated and documented, including computational resources (GPU/CPU), data requirements, model retraining needs, human expertise, costs, and implementation timelines;
vi. the appropriate approver evaluates each AI change request based on the results of the impact analysis and the justification provided. Evaluation criteria shall include expected improvements in model performance, fairness and bias implications, explainability requirements, regulatory compliance, cost–benefit considerations, and potential disruption to existing AI services;
vii. the disposition of each change request is determined as follows:

a. reject the change request where it is not approved due to unacceptable risk, limited benefit, or non-compliance with established standards and guidelines. Reasons for rejection shall be documented and communicated to the requester.
b. defer the change request where it is postponed for future consideration due to lower priority or dependencies on other system improvements or data availability; and
c. approve the change request where it is authorized for implementation, with assigned priority and conditions for deployment, and forward approved requests to the AI maintenance team for execution.

viii. the team develops a detailed modification plan based on the approved change request that shall include:
   a. identification of affected AI components (such as models, datasets, APIs, and pipelines);
   b. updates to model architecture or training strategy;
   c. how the changes will be executed and delivered; and
   d. the required testing activities and procedures for validating the changes.

ix. approved changes are implemented in a test environment, including model retraining, fine-tuning, or dataset updates as applicable;

x. any adverse effects identified during implementation shall be addressed and resolved before proceeding further;

xi. the updated AI system undergoes comprehensive testing and validation to ensure accuracy, security, stability, fairness, and compliance;

xii. testing includes technical performance evaluation, robustness testing, bias and fairness assessment, and where applicable, human oversight testing;

xiii. the modified AI system is deployed into production environment in a controlled manner to minimize operational risk;

xiv. continuous monitoring is conducted to detect performance drift, anomalies, unintended bias, security issues, and user impact; and

xv. where significant changes occur, appropriate user training, updated documentation, and stakeholder communication is provided to ensure safe and effective use of the system.

#### 3.1.2.8 Retirement Standards

Public institution intending to retire an AI system shall:

i. adhere to the Public Finance (Management of Public Property) Regulations GN. No. 373 of 2024;
ii. Develop and implement a detailed AI system retirement plan ensuring that all retirement activities are adequately addressed and completed;
iii. Obtain approval of the AI system retirement plan and final retirement outcomes from the institutional steering committee to confirm that all retirement activities, decommissioning processes, and transition arrangements have been adequately addressed; and
iv. ensure that the models are securely discontinued and historical data are archived or purged in accordance to Records and Archive Management guidelines.

## 3.2 THE GUIDELINES

### 3.2.1 General Guidelines

Public Institution shall ensure that:

i. Standards and Guidelines for Development, Acquisition, Operation and Maintenance of e-Government Applications are adhered when developing, acquiring, operating or using AI Systems;

ii. institutional ICT management documents are reviewed and updated where applicable to address AI governance, risk management and acceptable use;

iii. all AI solutions are recorded in the institutional ICT Asset Register prior to operational use;

iv. accountability for AI use is established through clear ownership, assigned responsibilities, approval authority, risk ownership, and monitoring duties;

v. AI governance involves the relevant stakeholders and not left solely to technical staff or vendors;

vi. existing AI systems in operation are identified and evaluated to ensure compliance with the established e-Government standards and guidelines;

vii. where an existing AI system presents non-compliance with the established e-Government Standards and Guidelines, the institution shall take appropriate corrective actions including remediation, restriction, replacement, suspension, or retirement of the AI system to ensure compliance;

viii. a competent team is formulated with appropriate competencies and expertise a per AI use case requirement to support the acquisition and implementation of AI system;

ix. personnel involved in the acquisition, implementation, operation, maintenance, use, and oversight of AI systems receive appropriate training and ongoing capacity building;

x. as far as practicable, the use of AI systems that operate as black-box or vendor lock-in solutions are avoided;

xi. AI impact re - assessment is performed whenever there is a material change in:

a) the purpose of the use case;

b) the role of AI;

c) the data involved;
d) the model, service, or platform used;
e) the users or affected persons;
f) the deployment environment;
g) the risk level; or
h) any other factor materially affecting the governance significance of the use case.

xii. Decision-support AI does not remove accountable human responsibility for consequential decisions or actions;

xiii. Decision-support AI is used in a manner that does not results in arbitrary, opaque, unfair, or unjustifiable action;

xiv. Where AI is used to support scoring, classification, prioritization, recommendation, profiling, or similar functions, the institution clearly defines:
a) the role of the AI use case;
b) the extent to which human judgment remains required;
c) the limits on reliance that may be placed on outputs; and
d) the escalation path where outputs are contested, unclear, or potentially harmful.

xv. material changes to an AI system, approved AI use case, data source, model, prompts, thresholds, integration, vendor service, hosting arrangement, or deployment environment are subject to testing and validation before continued use, corresponding with the risk and nature of the change;

xvi. document and maintain records of AI system assumptions, limitations, known constraints, foreseeable misuse scenarios, unsupported and prohibited uses, conditions under which the AI system shall not be used;

xvii. where an AI system forms part of a wider ICT system, or equipment the AI component remains identifiable, governable, testable, documented, and reviewable;

xviii. fallback, manual processing, appeal, complaint, correction, and escalation procedures are in place where the AI system affects public-service delivery, decisions, rights, interests, obligations, or user interaction; and

xix. train users on approved purpose, prohibited uses, known limitations, escalation points, human review requirements, data protection duties, and incident reporting.

### 3.2.2 Specific Guidelines

#### 3.2.2.1 Initiation

Public Institution shall:

i. develop an AI use case that includes:
   a) the problem to be addressed;
   b) the intended public value or institutional benefit;
   c) the responsible owner;
   d) the proposed AI solution, model, platform, or service;
   e) the expected users or affected persons;
   f) the main categories of data involved;
   g) the expected role of AI in the relevant process or decision; and
   h) any known legal, ethical, privacy, security or operational concerns.

ii. perform AI Impact Assessment by executing the following prescribed steps:
   a. define the assessment scope by setting boundaries, exclusions and depth of analysis;
   b. identify all relevant stakeholders such as individuals, group of individuals and societies affected by the AI Use Case;
   c. identify the potential benefit and negative impact for each stakeholder across various areas including but not limited to fairness and non-discrimination, privacy, safety and health, security, accountability, transparency and explainability, reliability, accessibility, human rights, financial consequences and environmental impact (**see Appendix I – Potential Benefit and Negative Impact,** and **Appendix II – Impact Areas)**;
   d. identify the predictable or known ways the AI system can fail and the impact on the relevant stakeholders as a result of such failures (**see Appendix III – Impacts of AI system failure**);
   e. identify the way in which the AI system can be reasonably foreseeable misused or intentionally abused and the impact on relevant stakeholders as a result of such reasonably foreseeable misuse or intended abuse (**see Appendix IV – Impacts of AI system misuse or abuse**);

f. assess the magnitude and likelihood of the identified impacts;
g. map each identified impact to the corresponding treatment (e.g mitigation measures);
h. determine the overall AI System Impact Level as either unacceptable, high impact, medium impact or low impact as depicted in **Appendix V – Overall AI System Impact Rating**. The overall AI system impact rating is determined by the highest impact level after application of controls; and
i. prepare an AI Impact Assessment Report.

iii. use the AI Impact Assessment results to justify the AI Use case, validate ethical alignment, and provide the evidence base necessary for the Institutional ICT Steering Committee approval;
iv. perform AI Risk Assessment using Institutional risk management procedures covering areas such as technical, security, data privacy, ethical, operational, legal and regulatory risks;
v. utilize the findings of the AI Impact Assessment as the foundational data for the AI Risk Assessment ensuring the technical risk mitigation are directly informed by the identified potential negative impacts to individuals, society, and organization;
vi. prepare an ICT project concept note for the AI use case in accordance with the Standard and Guidelines for Government ICT Project Implementation and submit it to the Institutional Steering Committee together with the AI Impact Assessment Report for review and approval; and
vii. submit the approved ICT project concept note and AI impact Assessment Report through the Government ICT Services Portal (GISP) for review, technical guidance and clearance to proceed.

#### 3.2.2.2 Design

Public institution shall ensure that:

i. AI systems are designed with appropriate technical and organizational safeguards to prevent, detect, and mitigate reasonably foreseeable misuse, unauthorized activities, and harmful outputs;

ii. the required resources are identified and documented, including computational resources (GPU/CPU), data requirements, human expertise, costs, and implementation timelines;
iii. data protection and privacy mechanisms such as anonymization, pseudonymization, tokenization or aggregation are considered during the design stage;
iv. establish a documented bias mitigation plan that identifies relevant sources of bias, affected groups, fairness measures and acceptance thresholds, and applies appropriate pre-processing, in-processing and post-processing techniques based on the intended use and level of risk;
v. security requirements such as authentication, authorization, encryption, logging, rate limiting and segregation controls are specified for AI components;
vi. Robustness testing plan includes malformed inputs, edge cases, load conditions and adversarial conditions; and
vii. the oversight design systematically account and manage human reviewer workload so that human review of AI-generated outputs remains thorough, effective, and meaningful, without being compromised by excessive volume or complexity.

#### 3.2.2.3 Acquisition Guidelines

**a. Guidelines for Open-Source AI Acquisition**

Public institution intending to acquire open-source AI solution shall ensure that:

i. the AI solution is capable of being customized to fit the intended use case;
ii. understands the model's purpose, functionality, training approach, data sources where available, input requirements, limitations, and known risks before adopt or deploy open-source models;
iii. a clear understanding of the AI solution by incorporating evaluation criteria that assess the algorithms and models used, including variable selection, underlying AI techniques such as supervised, unsupervised, reinforcement learning, model limitations, and the source and nature of training data, and may require independent audits of the algorithms where appropriate;
iv. comprehensive technical documentation are available and maintained that include but not limited to:

a) a general description of the AI solution including its intended purpose;
b) usage instructions;
c) technical assumptions about its deployment and operation such as run-time environment, related software and hardware capabilities, and assumptions made on data;
d) technical limitations such as acceptable error rates, accuracy, reliability, robustness; and
e) monitoring capabilities and functions that allow users or operators to influence the system operation.

v. a thorough security assessment of the open-source AI component is conducted before deployment, including static code analysis and penetration testing;
vi. licenses of all open-source AI components used are identified, documented, understood, and regularly monitored to ensure compliance and prevent conflicts with government security or confidentiality requirements; and
vii. adopts open-source AI solutions that demonstrate active and sustainable community and viable means of obtaining ongoing technical support.

**b. Guidelines for Commercial Off-the-Shelf (COTS) AI Solutions**

Public Institution shall ensure that:

i. proper data governance mechanisms, including rules on access, sharing, storage, and consent, are in place from the beginning;
ii. the COTS AI systems are capable of being customized to fit the intended use case;
iii. clearly set out who owns what in every AI contract, including ownership of models, code, data and outputs;
iv. contracts shall prohibit vendors from using government data that are not publicly accessible to train their own commercial AI products without the institution's written consent, and shall include clear rules on how data is handled, stored, and eventually deleted;
v. AI providers support knowledge transfer, staff training, and long-term risk monitoring; and

vi. track the performance of COTS AI system to make sure it performs the intended purpose and continues to deliver reliable results.

### c. Guidelines for Outsourced AI development

Public institution shall ensure that:

i. the internal institutional team is involved throughout the development of the AI system;

ii. vendors are evaluated based on more than cost, including their public sector AI experience, governance practices and certifications, technical team expertise, data security measures, access to source code and documentation, audit rights, financial stability, business continuity plans, and verified references from government projects;

iii. the vendor provides version control for all code, models, and datasets; regular progress reporting; continuous testing against agreed acceptance criteria; and immediate reporting of any scope changes, data issues, or new risks for approval;

iv. maintenance agreements include service levels for system availability and response times, model retraining, system improvements, security patching and vulnerability management;

v. knowledge transfer plan and obligations are included in all contracts, requiring vendors to train institutional staff to operate, monitor, and troubleshoot the system, provide full technical and operational documentation, and implement a phased handover plan with clear milestones to reduce reliance on the vendor over time;

vi. vendor contract clearly specifies the handover of all source code, trained models, configuration files, and documentation upon termination, and requires the return or secure deletion of all institutional data; and

vii. vendors implement and maintain appropriate information security and privacy controls to protect the confidentiality, integrity, availability, and privacy of information processed on behalf of the institution.

### d. Guidelines for AI as a Service

Public institution shall ensure that:

i. before acquiring AI as a Service assesses:

a) the purpose of use;
b) the type of information involved;
c) the provider or platform dependency;
d) the security, privacy, confidentiality, and continuity implications;
e) the level of transparency, explainability and control available; and
f) any legal, contractual, or operational constraint affecting proper use.

ii. AI as a service is used where the institution has the team to adequately govern, supervise, explain, restrict, or discontinue its use;

iii. personal data used in AI as a Service comply with all applicable privacy and data protection laws and regulations including Personal Data Protection Act, Cap.44, R.E.2023;

iv. management and use of Confidential information into AI as a Service shall comply with the National Security Act, Cap.47, R.E.2023, Records and Archives Management Act, Cap 309, R.E.2023, institutional information classification policy as well as other applicable data protection laws and regulations; and

v. approval is obtained from e-Government Authority prior to acquisition or usage of AI as a Service.

### e. Guidelines for Due Diligence of a third-party AI System

Public institution shall ensure that:

i. vendor due diligence is proportionate to the risk level of the AI use case, the criticality of the public function involved, the sensitivity of the data involved, and the level of dependence on the provider;

ii. due diligence considers:

a) the identity, capability, and role of the provider;
b) the nature of the model, platform, or service being provided;
c) the provider's governance, security, privacy, confidentiality, and continuity arrangements;

d) the level of transparency, documentation, and traceability availability;
e) the provider's incident handling, support, and change management arrangements;
f) any subcontracting, hosting, or external dependency arrangements;
g) legal, operational, interoperability, and exit implications; and
h) any material limitation, condition, or exclusion affecting lawful and accountable use.

iii. where due diligence identifies material concerns relating to lawfulness, privacy, confidentiality, security, continuity, explainability, documentation, or dependency, the institution shall not proceed until the concerns are resolved.

#### 3.2.2.4 Development

##### a. Data Management and Preparation Process

Public institution shall:

i. establish controlled and lawful procedures for data collection, data quality management, usage, sharing, storage, archival, retention and disposal, while ensuring full traceability and accountability throughout the entire data lifecycle;
ii. implement standardized processes for data cleaning, transformation, profiling, validation, integration, and feature engineering to ensure that all data used in AI systems is accurate, consistent, complete, relevant, and interoperable;
iii. maintain documentation describing dataset, data sources, preparation methods, quality assessment, transformation and preprocessing activities;
iv. ensure that datasets are contextual, representative, unbiased, and properly separated into training, validation, and testing datasets to preserve dataset integrity and support objective AI model development and evaluation;
v. verify that training, validation, and test datasets remain properly partitioned, free from duplicate or overlapping records, and protected from data leakage by ensuring that preprocessing transformations are fitted solely on the training dataset;
vi. use trained data labelers and standardized data labelling procedures to label data consistently and correctly, while reducing errors and bias in the labelling process; and

vii. retain data sources and datasets used by AI systems in accordance with approved data retention requirements and shall securely archive or dispose of the data upon expiry of the applicable retention period.

**b. Model Training**

Public institution shall:

i. have a configured adequate computational infrastructure to support AI model training, aligned with data scale and algorithm complexity before training;
ii. implement robust security controls to safeguard the training environment, including access management, encryption, and continuous vulnerability monitoring;
iii. manage risks associated with third-party tools and libraries, ensuring they meet security, legal, and institutional compliance requirements;
iv. define suppression rules and implement grounding mechanisms to prevent hallucinated, harmful, or out-of-domain outputs;
v. evaluate model reuse feasibility before building from scratch and perform domain-specific fine-tuning to match the target environment;
vi. confirm that the chosen model meets all required thresholds for accuracy, bias control, and safety before giving final approval, and keep written records explaining why that model was selected over the others to support transparency and future reviews;
vii. keep an organized record of every model and dataset version, documenting what changed in the training data, settings, and code, to ensure that any past version of the model can be recreated or reviewed; and
viii. ensure that before fine-tuning pre-trained models the underlying data, behavior and risks are well understood and controlled.

### 3.2.2.5 Testing and Validation Guidelines

Public Institution shall:

i. test and validate acquired AI systems using representative local data, local users, local workflows, and local operating conditions, where applicable to verify their suitability for the intended deployment context;

ii. define output review criteria for AI systems where applicable, covering factual accuracy, traceability, completeness, sensitivity, confidentiality, harmful outputs, public communication risks, and human review requirements;
iii. confirm that bias is not reintroduced through proxy features after protected attributes have been excluded from model inputs;
iv. evaluate AI systems for bias using independent datasets that are, as far as practicable, free from the bias being assessed;
v. test the effectiveness of human oversight under realistic operational workloads to ensure that human review remains meaningful and effective; and
vi. re-test or re-validate AI systems following material changes, including model updates, retraining, data source changes, data or concept drift, performance degradation, integration changes, security incidents, legal or regulatory changes, or changes to the intended purpose.

#### 3.2.2.6 Deployment Guidelines

Public Institution shall ensure that:

i. deployment authorization request includes approved AI use case, risk classification, deployment scope, residual risks, deployment conditions, testing and validation results, deployment readiness assessment, deployment plan, monitoring plan, and fallback plan;
ii. AI hosting environment is assessed, configured, tested, and confirmed as ready before deployment, including network connectivity, access controls, system integrations, logging, monitoring, backup and recovery arrangements, support arrangements, and production data pipelines required for secure, reliable, and controlled operation;
iii. apply appropriate safeguards, including access, capability, usage, content, data boundary, throttling, and human approval controls, where necessary;
iv. AI systems are formally handed over to operational team before production use, supported by system documentation, user guidance, deployment restrictions, monitoring procedures, fallback procedures, vendor contacts if applicable, support arrangements, known limitations, unresolved issues, and user responsibilities; and

v. AI systems, including those deployed on shared infrastructure, are hosted only in approved environments with appropriate security, access control, isolation, logging, monitoring, backup, recovery, and support arrangements.

#### 3.2.2.7 Operation and Monitoring Guidelines

##### a. Guidelines for AI Operation

Public Institution shall ensure that:

i. access to the AI model, infrastructure and data is restricted to authorized users only, in adherence to the principle of role-based access control;
ii. regular backup of AI model and data is performed to prevent loss due to unforeseen circumstances such as hardware failure;
iii. the AI system undergoes security assessment at least semi-annually and has in place a documented process for all AI operations, such as user and access management; and
iv. audit trail is established to monitor and record all activities related to inputs, outputs, decisions made by AI, system changes and updates.

##### b. Guidelines for AI Maintenance

Public Institution shall ensure that:

i. has in place documented procedures to guide any changes to AI systems.
ii. any changes, such as, model retraining, bugs and errors fixing, AI model patching and upgrading are undertaken in accordance with institutional change management process and are properly documented;
iii. AI system changes are tested prior to deployment, and clearly documented for future reference;
iv. changes made to AI system can be rolled back if required;
v. AI system performance and outputs are revalidated following the implementation of any release, update, or modification;
vi. AI Systems are subject to periodic reviews and revalidation to ensure alignment with updated institutional policies, laws and operational needs;
vii. records of AI changes, reviews, decisions, and approvals are maintained;

viii. AI systems, which require licenses, have a valid maintenance license for patches and upgrades, including presence of a clear Service Level Agreements (SLA); and

ix. AI models are updated regularly to mitigate risks arising from outdated data, unfair or skewed outcomes/ decisions.

### c. Guidelines for AI Monitoring

Public institution shall ensure that

i. AI system or model performance criteria/parameters are identified, regularly monitored, and tracked throughout the AI operational lifecycle, including accuracy and consistency of outputs, system performance, emerging risks and control weaknesses;

ii. monitoring and tracking of AI Infrastructure performance including CPU, GPU, memory, logs, services, processes, and network usage is performed to avoid bottlenecks;

iii. appropriate mechanisms are established to detect anomalies and AI system failures, monitor AI outputs and decisions, identify bias, model drift, or harmful behavior, and other related security threats;

iv. monitoring of safety including monitoring inputs and outputs for abuse, misuse, information disclosure and any sense of harm is performed;

v. monitoring of AI system reliability including error rates, fault detection, recovery, redundancy and failover mechanisms is performed;

vi. periodic assessments are conducted post-deployment to verify that fairness, transparency, explainability, security and robustness requirements continue to be met;

vii. monitoring of human oversight, including assessing the effectiveness of human oversight and control measures, analyzing usage metrics to identify operational challenges, friction points, and opportunities to improve the overall effectiveness and outcomes of human oversight;

viii. deviations from established AI performance criteria are identified and analyzed to determine their underlying causes and potential impacts, and that recommendations are provided on necessary adjustments to the AI system based on the established performance criteria; and

ix. feedback mechanism is implemented to allow users to report unfair decisions, harmful outputs and accessibility challenges.

### d. Incident Management and Escalation

Public institution shall ensure that:

i. AI-related Incidents are detected, classified, assessed, recorded, and managed through approved institutional incident management procedures;

ii. AI system incidents are promptly assessed and investigated to determine root causes and evaluate their operational, ethical, legal, and security impacts, taking into consideration factors such as data quality issues, model errors or drift, failures in human oversight, and security vulnerabilities;

iii. AI incidents are timely and appropriately escalated in accordance with institutional incident management procedures and reported to e-GA through e-Government Security Operation Center (eGSOC), depending on their nature and severity;

iv. AI systems are safely restored, revalidated prior to resuming operation, and appropriate corrective measures are implemented; and

v. a documentation of AI incidents is maintained that shall include, the nature of the incident, affected systems or users, actions taken, escalation made, lessons learnt, and follow-up measures.

#### 3.2.2.8 Retirement Guidelines

Public institution shall ensure that:

i. a documented retirement plan is established prior to AI system decommissioning. The retirement plan shall include:
   a. retirement decisions;
   b. scope;
   c. roles and responsibilities;
   d. implementation timelines;
   e. transition arrangements;
   f. data handling procedures;
   g. access revocation procedures;
   h. infrastructure decommissioning activities;
   i. record retention requirements;
   j. risks and mitigation measures; and

k. post-retirement monitoring arrangements.

ii. AI systems are retired where they no longer meet operational requirements, present unacceptable risks, experience significant performance degradation, become technologically obsolete, fail to comply with updated legal or regulatory requirements, or are replaced by improved solutions;

iii. institutional change management procedures are followed before implementing AI system retirement;

iv. AI systems are retired in a controlled and accountable manner that reduce operational disruption, protect data and systems, preserve auditability, prevents unauthorized access or reuse following decommissioning, secure disposal of storage media where applicable and complies with applicable legal and regulatory requirements; and

v. records relating to AI system are retained in accordance with institutional and applicable regulatory retention requirements.

### 3.2.3 Guidelines on Prohibited AI Use Cases

Due to the current limitations of AI technologies and their potential ethical, legal, and social implications, certain AI use cases are considered inappropriate for use within the public sector and are prohibited. These include, but are not limited to:

i. fully automated decision-making AI without human control involving significant outcomes, particularly in areas affecting an individual's health, safety, rights, education, employment, or access to public services. These systems may lead to inaccurate outcomes, lack of human judgment, reduced accountability, and limited opportunities for individuals to challenge or appeal decisions;

ii. AI systems that evaluate, rank, or assign scores to individuals based on their behaviors, personal characteristics, financial status, online activities, social interactions, or historical data in order to determine their eligibility, priority, or level of access to government or public services and benefits. This practice may lead to discrimination, exclusion, unfair treatment, loss of privacy, and unequal access to essential public services, especially where the scoring criteria are unclear, biased, or lack transparency and accountability;

iii. AI enabled Biometric systems that classify individuals based on sensitive personal characteristics such as race, religion, or health status. These systems may result in discrimination, profiling, privacy violations, stigmatization, and misuse of personal data, particularly where individuals are categorized without consent or where the technology lacks accuracy, fairness, or legal safeguards;

iv. AI systems that analyze or predict emotions for decisions related to employment, education, or credit access. These systems may lack scientific reliability and can produce inaccurate or biased outcomes that negatively affect individuals' rights, opportunities, dignity, and equal treatment; and

v. AI Systems that secretly or unfairly influence the behavior or decisions of vulnerable individuals or groups. These systems may exploit cognitive, emotional, or behavioral vulnerabilities to influence political opinions, online behavior, or access to services that may undermine individual autonomy, informed decision-making, human dignity, and public trust.

### 3.2.4 Guidelines for Responsible Use of Generative AI

For responsible use of Generative AI, a person working with Government Information shall ensure that:

i. understands the various risks associated with the use of generative AI tools including but not limited to:
   f. Outdated information: Generative AI tools may have been trained on historical data with fixed cut-off dates and may not have access to current information;
   g. Incompleteness: AI generated content may omit relevant arguments and issues or overlook essential points;
   h. Inaccuracy or fabrications: Generative AI tools can invent facts and confidently present them as true; and
   i. Bias: Generative AI tools may embed biases by over or under representing certain demographics.

ii. protects and maintains privacy and confidentiality by not putting confidential, personal or copyrighted information into generative AI tools;

iii. sensitive ICT system details including system documentation, configuration files, source codes, passwords, or system access credentials are not entered into AI systems that are not owned by the Government;

iv. generative AI outputs are reviewed before being relied upon for official use to ensure accuracy, fairness and non-discrimination;

v. remain responsible and accountable for the content created, shared or used;

vi. understands the capabilities and limitations of generative AI tools used in assisting undertaking of Government operations;

vii. generative AI tools are not used to create, alter, or communicate official content in a misleading, deceptive, or unauthorized manner;

viii. uses the "opt-out" feature where possible to ensure that prompts are not used to train or further develop the generative AI tool;

ix. never delegate to an AI system any task that you could not personally perform, critically evaluate, or take full responsibility for;

x. forms their own views before seeking ideas or recommendations from generative AI tools; and

xi. informs users when audio or visual messages addressed to them are generated by AI.

## 4 IMPLEMENTATION, REVIEW AND ENFORCEMENT

This document shall be:

4.1 Effective upon being reviewed and approved by the Authority Board of Directors.

4.2 Subjected to review at least once every three years or whenever necessary changes are needed.

4.3 Consistently complied with, any exceptions to its application must duly be authorized.

## 5 RELATED DOCUMENTS

5.1 The e-Government Act, Cap.273, R.E.2023;

5.2 The e-Government (General) Regulations, GN. No. 75 of 2020 as amended by GN No. 375 of 2025;

5.3 National ICT Policy, 2016;

5.4 e-Government Infrastructure Architecture Standards and Technical guidelines (eGA/EXT/IRA/001).

5.5 e-Government Application Architecture -Standards and Technical Guidelines (eGA/EXT/APA/001).

5.6 Tanzania Development Vision 2050 (Dira 2050);

5.7 Tanzania Digital Economy Strategic Framework 2024 – 2034;

5.8 Tanzania e-Government Strategy, 2022;

5.9 e-Government Technology Roadmap, 2024;

5.10 UNESCO Tanzania AI Readiness Assessment Report, 2025;

5.11 The Personal Data Protection Act, Cap 44, R.E.2023;

5.12 UNESCO Recommendation on the Ethics of Artificial Intelligence, 2021;

5.13 OECD AI Principles, 2024;

5.14 AU Continental AI Strategy, 2024;

5.15 e-Government Security Operations Guideline (eGA/EXT/ISA/004);

5.16 e-Government Guideline, 2017;

5.17 Standards and Guidelines for Development, Acquisition, Operation and Maintenance of e-Government Applications (eGA/EXT/APA/005);

5.18 Procedures for Government ICT Project Clearance, Monitoring and Closure (eGA/EXT/BSA/009);

5.19 Standards and Guidelines for Government ICT Project Management Implementation (eGA/EXT/BSA/008);

5.20 Quality Assurance Compliance Guidelines for e-Government Applications (eGA/EXT/APA/007);

5.21 e-Service Sustainability Framework (eGA/EXT/BSA/006)

5.22 e-Government Security Architecture Standards and Technical Guidelines (eGA/EXT/ISA/001);

5.23 e-Government Emerging Technology Framework (eGA/EXT/AVS/008);

5.24 NIST AI Risk Management Framework (AI RMF 1.0), 2023;

5.25 ISO/IEC 42001:2023 Information Technology – Artificial Intelligence – Management System; and

5.26 ISO/IEC 22989:2022 Information Technology — Artificial Intelligence — Artificial Intelligence Concepts and Terminology.

## 6 DOCUMENT CONTROL

| Version | Name | Comment | Date |
|---|---|---|---|
| Ver. 1.0 | e-GA | Creation of Document | August 2026 |

# APPENDIX

## Appendix I: Potential Benefit and Negative Impacts

For each identified stakeholder, identify potential benefit and negative impact.

| Areas | Potential Benefit | Potential Negative Impact |
|---|---|---|
| Fairness and Non-discrimination | | |
| Privacy | | |
| Safety and Health | | |
| Security | | |
| Accountability | | |
| Transparency and Explainability | | |
| Reliability | | |
| Accessibility | | |
| Human rights | | |
| Financial consequences | | |

## Appendix II: Impact Areas

| Areas | Examples of Considerations |
|---|---|
| Fairness and non-discrimination | • The impact to the relevant stakeholder if the AI system performance is worse for any particular population groups that the stakeholder associates themselves with.<br>• The impact to the relevant stakeholder if the AI system represents the interested party in the ways that discriminate against them based on their associated population groups. |

| Areas | Examples of Considerations |
|---|---|
| Privacy | • The impact of the relevant stakeholders if personal data has been collected for the purpose of AI system development or use (e.g model training).<br>• The impact to the relevant stakeholder resulting from a privacy breach related to the processing of personal data performed for the development or use of the AI System. |
| Security & Safety | • The impact to the relevant stakeholder if the AI system performs unsafely.<br>• The impact to the relevant stakeholder if the event of changes to the AI system including as a result of identifying issues or improvements to the AI system performance.<br>• The impact to the relevant stakeholder due to known false positives, false negatives and other known failures. |
| Accountability | • The potential for the use of the AI system to result in an impact to the legal position, life opportunities or human rights.<br>• The potential for the use of the AI system to cause physical or psychological harm to the stakeholder.<br>• Impacts to the relevant stakeholder in the event the AI system does not adequately fulfill the purpose it was deployed for.<br>• Impacts to the relevant stakeholder as a result to data used to develop the AI System.<br>• Impacts to the relevant stakeholder as a result of not applying human oversight to the AI System and its outputs. |
| Transparency | • The impact to the relevant stakeholders if there is not enough information about the capabilities of the AI system, its constraints and limitations. |

| Areas | Examples of Considerations |
|---|---|
| | • The impact of the relevant stakeholders if they are unaware that they are interacting with the AI system. |
| Explainability | • The impact to the relevant stakeholder if there is not enough information available to make informed decisions based on the outputs of the AI System.<br>• The impact of explainability on the relevant stakeholders during design, development, testing, validation and deployment stages of the AI System life cycle.<br>• The impact of explainability on the trust of the relevant stakeholders in the AI System.<br>• The impact of explainability on the ability relevant stakeholders to dispute the outputs of an AI system. |
| Reliability | • The impact of the relevant stakeholder if the AI System does not meet stated performance requirements during its operation stage.<br>• The impact to the relevant stakeholder if the event an update to the AI System, data drift or concept drift causes its performance to fail meet requirements.<br>• The impact to the relevant stakeholder arising from predictable failures such as false positives or false negatives.<br>• The impact to the relevant stakeholder if the AI system or its outputs are compromised by unauthorized parties resulting in failure to meet stated performance requirements. |
| Accessibility | • Accessibility aspects for the relevant stakeholder so they are not excluded due to disabilities or lack of digital skills.<br>• The impact on the relevant stakeholder if the AI System disproportionately affects children, persons with disabilities, |

| Areas | Examples of Considerations |
|---|---|
| | older persons, or other vulnerable groups by failing to provide appropriate safeguards and reasonable accommodations. |
| Human rights | • The impact on the relevant stakeholder if the AI System infringes, restricts, or otherwise adversely affects their human rights during its operation stage.<br>• The impact on the relevant stakeholder arising from predictable AI errors, including false positives, false negatives, or inaccurate outputs, that may result in the denial, restriction, or impairment of their human rights.<br>• The impact on the relevant stakeholder if the AI System lacks appropriate human oversight, resulting in decisions that adversely affect fundamental rights without meaningful human intervention or review.<br>• The impact on the relevant stakeholder if the AI System limits or adversely affects access to essential public services, employment, education, healthcare, justice, or other rights due to inaccurate, biased, or automated decision-making.<br>• The impact on the relevant stakeholder if the AI System enables unlawful surveillance, profiling, or monitoring that interferes with the rights to privacy, freedom of expression, freedom of association, or other fundamental freedoms. |
| Financial consequences | • The financial impact on the relevant stakeholder if the AI System does not meet stated performance requirements during its operation stage.<br>• The financial impact on the relevant stakeholder arising from predictable AI errors, including false positives, false negatives, |

| Areas | Examples of Considerations |
|---|---|
| | or inaccurate outputs that result in financial loss, incorrect transactions, or missed opportunities.<br>• The financial impact on the relevant stakeholder if the AI System or its outputs are compromised by unauthorized parties, resulting in fraud, theft, financial loss, or unauthorized financial transactions.<br>• The financial impact on the relevant stakeholder if the AI System becomes unavailable, experiences prolonged downtime, or suffers service disruptions that affect business continuity or revenue generation.<br>• The financial impact on the relevant stakeholder arising from non-compliance with applicable legal, regulatory, contractual, or financial obligations attributable to the AI System, including fines, penalties, compensation, or litigation costs. |
| Environmental Impact | • The environmental impact arising from the energy consumption of the AI System throughout its lifecycle, including development, training, deployment, operation, maintenance, and decommissioning.<br>• The environmental impact arising from the computational resources, infrastructure, and hardware required to develop, train, deploy, operate, and maintain the AI System.<br>• The environmental impact arising from the AI System's consumption of natural resources, including electricity, water, and critical materials used in computing infrastructure.<br>• The environmental impact arising from the storage, processing, and retention of data beyond what is necessary for the intended purpose of the AI System. |

| Areas | Examples of Considerations |
|---|---|
| | • The environmental impact arising from the procurement, operation, maintenance, and disposal of hardware supporting the AI System, including electronic waste and resource depletion.<br>• The environmental impact if the AI System supports or promotes environmentally sustainable practices, including improving resource efficiency, reducing waste, minimizing emissions, or optimizing the use of energy, water, or other natural resources.<br>• The environmental impact arising from the failure, misuse, or compromise of the AI System that results in environmental harm, increased pollution, inefficient resource utilization, or disruption of environmentally critical operations.<br>• The environmental impact arising from the decommissioning or replacement of the AI System, including the management of electronic waste, hardware disposal, and opportunities for reuse or recycling. |

## Appendix III: Impacts of AI system failure

| S/N | Failure description | Impact resulting from potential failure |
|---|---|---|
| 1. | | |
| 2. | | |
| 3. | | |

## Appendix IV: Impacts of AI system misuse or abuse

| S/N | Misuse or abuse description | Impacts resulting from potential misuse or abuse |
|---|---|---|
| 1. | | |
| 2. | | |
| 3. | | |
| 4. | | |

## Appendix V: Overall AI System Impact Rating

| Highest Impact Level | Overall Impact Rating |
|---|---|
| **Critical** | Unacceptable |
| **High** | High Impact |
| **Medium** | Medium Impact |
| **Low** | Low Impact |

President's Office,
e-Government Authority,
Mawasiliano Street,
P.O.Box 2833, DODOMA.

Website : www.ega.go.tz
Email: barua@ega.go.tz

"Building a Digital Government"