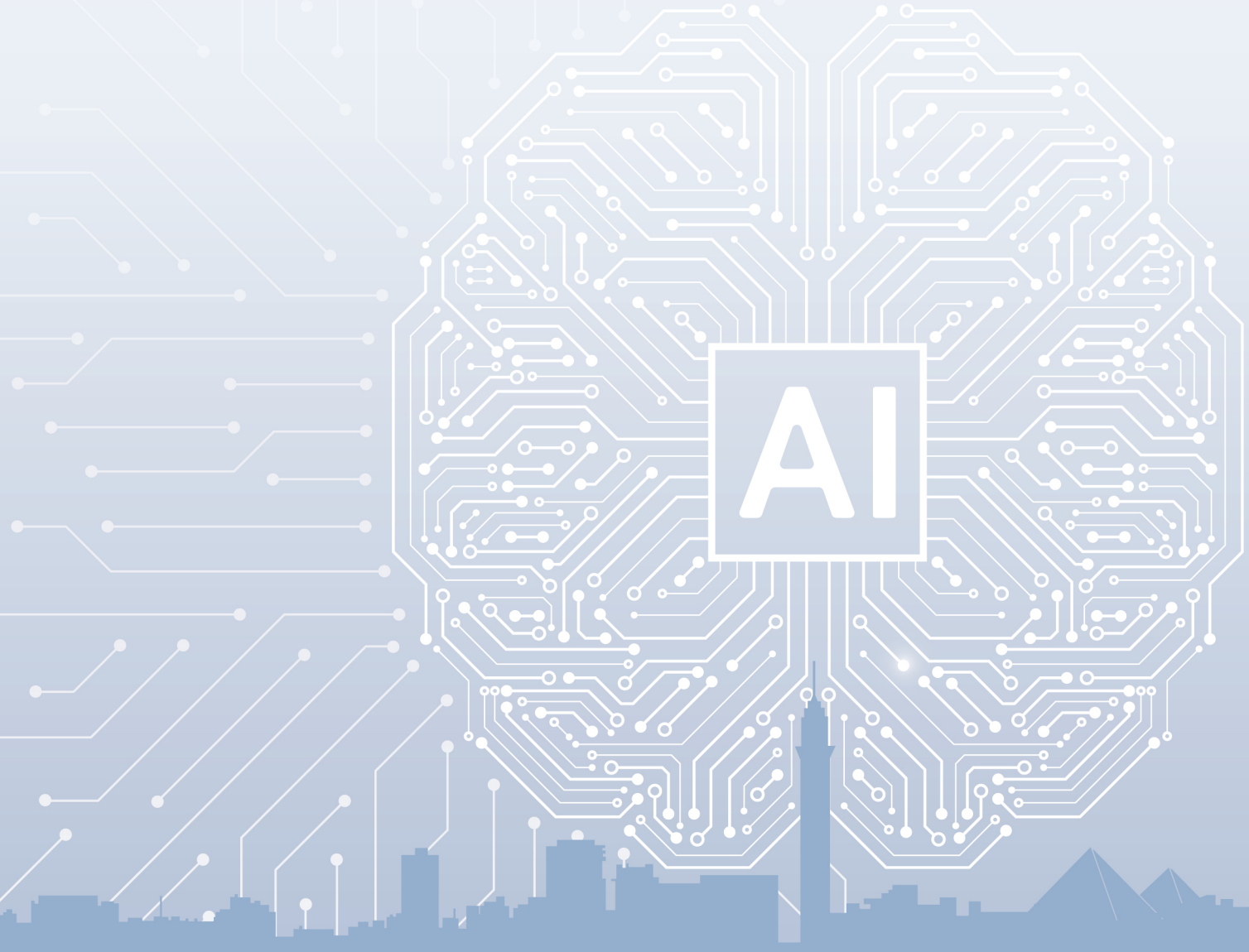
National Council for Artificial Intelligence, Quantum Computing, and Emerging Technologies

Egyptian Center for Responsible AI

Arab Republic of Egypt
National Council for Artificial Intelligence,
Quantum Computing and Emerging Technologies
Egyptian Center for Responsible AI

EGYPT NATIONAL GUIDELINES

FOR TRUSTWORTHY AND RESPONSIBLE AI



March 2026

Egypt National Guidelines for Trustworthy and Responsible AI

Edition:	2.0
Status:	Approved for Publication
Date:	March 2026
Prepared by:	The Egyptian Center for Responsible Artificial Intelligence (ECRAI) Under the auspices of the National Council for Artificial Intelligence, Quantum Computing, and Emerging Technologies

Document Control & Administration

Document Information

Attribute	Details
Document Title	Egypt National Guidelines for Trustworthy and Responsible AI
Document ID	EG-NCAI-GOV-FW-2026-03
Classification	PUBLIC
Owner	National Council for Artificial Intelligence, Quantum Computing, and Emerging Technologies
Custodian	Egyptian Center for Responsible AI (ECRAI)
Applicability	All

Version Control

Version	Date	Author / Contributor	Description of Change
0.1	Aug 2025	Drafting Committee (ECRAI)	Initial structural draft and comparative analysis
0.8	Oct 2025	Drafting Committee (ECRAI)	Governance Philosophy and Main Directions
1.0	Jan 2026	Drafting Committee (ECRAI)	Final Version for Public Consultation
1.5	Feb 2026	Government, Private Sector and International Organizations	Integration of feedback from the Public Consultation
1.8	Feb 2026	Drafting Committee (ECRAI)	Harmonization and proofreading
2.0	March 2026	NCAI	Final Release

Authorization and Positioning

This document constitutes **Product 1** of the **National AI Governance Framework.**

It serves as the **Operational Manual** for the implementation of Trustworthy and Responsible AI practices in Egypt. It is technically distinct from, yet compliant with, the **Guide to the National AI Governance Framework** (The Policy Architect).

It has been reviewed for legal and technical consistency and authorized for release by:

H.E. Eng. Raafat Hindy

Minister of Communications and Information Technology
Chairman, National Council for Artificial Intelligence, Quantum Computing, and Emerging Technologies
(Signature Placeholder)

Date: March 25, 2026

Foreword

Artificial Intelligence is rapidly reshaping public services, businesses, and societies around the world. For Egypt, AI represents a historic opportunity to accelerate national development, improve government efficiency, and enhance the quality of life for citizens. But realizing this potential requires more than technological progress, it requires clear Governance to ensure that AI is used safely, ethically, and in a manner aligned with our national values.

These AI Guidelines for Trustworthy and Responsible AI provide a foundational, practical reference for all institutions—public and private—seeking to deploy AI responsibly. They translate the principles of our National Artificial Intelligence Strategy into actionable Governance measures that support innovation while safeguarding citizens' rights, promoting transparency, and ensuring trust in digital systems.

The Guidelines draw upon leading international practices and global standards, adapted to Egypt's context, legal environment, and development priorities. They are the result of collective effort, extensive consultation, and a shared vision among government entities, industry experts, academia, and civil society.

This document emphasizes a balanced Governance model: centralized policy-setting at the national level, combined with distributed sector-specific oversight to ensure that AI applications remain context-sensitive, fair, and accountable. It also highlights the importance of readiness—through institutional structures, trained workforces, and robust procedures—to ensure that AI systems are deployed responsibly across Egypt.

These Guidelines are not a static policy. As AI technologies evolve, the Governance approach must evolve as well. The Ministry, together with the National Council for AI, remains committed to updating and strengthening these Guidelines to keep pace with technological advancements and societal expectations.

I call upon all stakeholders—government agencies, private enterprises, developers, and communities—to adopt these Guidelines and work together to ensure that AI becomes a trusted enabler of Egypt's prosperity, resilience, and digital future.

Raafat Hindy

Minister of Communications and Information Technology
Chairman, National Council for Artificial Intelligence, Quantum Computing, and Emerging Technologies

Executive Summary

The National Guidelines for Trustworthy and Responsible AI (the Guidelines) establish a national reference for the responsible development, deployment, and oversight of Artificial Intelligence systems across the public and private sectors. These Guidelines aim to ensure that AI is used safely, transparently, and ethically, while supporting innovation and digital transformation in alignment with Egypt's Vision 2030 and the National AI Strategy.

Purpose of the Guidelines

While the **Guide to the National AI Governance Framework** defines **WHAT** should be governed (Risk Tiers and Mandates), these **Guidelines** define **HOW** to comply. This document provides the methodologies, metrics, and checklists required to translate the high-level Ethical Principles into Governance realities. Amongst others, it is particularly designed for Data Scientists, Compliance Officers, and Developers.

These Guidelines provide actionable directions for organizations on how to manage AI systems responsibly. The Guidelines seek to:

- Protect individuals' rights and societal well-being
- Ensure ethical and trustworthy use of AI
- Strengthen institutional accountability and transparency
- Promote innovation grounded in safety and sound governance
- Align Egypt with leading international standards and practices

To Whom the Guidelines Apply?

The Guidelines apply to all AI Actors in Egypt, including:

- Government entities (G)
- Enterprises and private-sector developers (E)
- Community actors and individuals (C)

Each Actor may serve as a Provider/Developer (PD) or Beneficiary/User (BU) of AI systems. The Guidelines apply across the full AI lifecycle—from design and data preparation to deployment, monitoring, and retirement.

AI-Level Assessment

These Guidelines introduce two types of AI-level assessment:

- AI-Institution Readiness Level: A comprehensive Self Assessment checklist representing an Internal Audit mechanism helping different organizations (G, E, or C) to assess and elevate their AI readiness level.
- Responsible AI-System Readiness Level designed to:
  - Assure "No Harm" by mitigating and combating potential risks.
  - Be ready when an AI law or decree is enacted
  - Ensure Compliance with international standards and norms

Key Components of the Guidelines

The document is organized around four core pillars that collectively define Trustworthy and Responsible AI Governance in Egypt:

1. Institutional Governance

Establishes the internal structures and procedures organizations must implement, including risk management policies, data governance, human oversight mechanisms, and workforce readiness. Institutions must adopt clarity in roles (such as Chief AI Officers), maintain audit mechanisms, and ensure continuous alignment with national regulations.

2. AI Lifecycle Governance

Provides guidance for responsible AI system development and deployment, including:

- **Design & Development:** embedding trustworthiness from the outset
- **Pre-deployment Assessment:** testing, evaluating, verifying, and validating AI systems (TEVV)
- **Post-deployment Monitoring & Audit:** ongoing oversight to identify and address emerging risks

3. Stakeholder Engagement

Encourages transparent communication with users and affected groups, promotes inclusivity in AI adoption, and fosters collaboration across government, industry, academia, and civil society.

4. Society & Sustainability

Addresses societal impact, cultural alignment, environmental considerations, and safeguards for vulnerable groups. It introduces concepts such as Frugal AI to reduce computational cost and environmental footprint.

Conclusion

These Guidelines position Egypt to responsibly harness the transformative power of AI while protecting society and strengthening institutional integrity. Their adoption will help cultivate a trusted, ethical, and forward-looking AI ecosystem that supports national development and ensures that AI technologies contribute positively to Egypt's future.

Table of Contents

Part I: Concepts and Foundations

1. Introduction

AI has been widely present in our daily life, from an analysis for the release of credit to a simple printing of a post on a citizen's social network timeline. AI development has been tremendously accelerating including Generative AI and Agentic AI where systems can autonomously perform multi-step tasks, interact with software environments, and make decisions with minimal human input.

The tremendous powerful capabilities of AI can be harnessed for tremendous benefit or severe harm. It is then the responsibility of global and national authorities to achieve the right balance between AI innovation and governance. In this sense, responsible development and use of AI is not about limiting innovation, but rather about ensuring its progress while balancing both social and technology aspects towards being human-centered, reliable, and aligned with social values. Further, responsible AI is not one-time task or rigid framework but an ongoing process involving all stakeholders across the whole journey starting with design and ending with deployment and monitoring.

AI systems consist not only of components using AI technology but can include a variety of technologies. Objectives such as safety, security, privacy and environmental impact should be managed holistically and not separately for AI and the other components of the system. Integration of the AI management system with generic or sector specific management system standards for relevant topics is therefore essential for responsible development and use of an AI system.

For Egypt, AI presents an unparalleled opportunity to accelerate national development, enhance public services, and position the country as a regional leader in responsible digital innovation. However, realizing these benefits demands a comprehensive Governance Framework that ensures AI systems are trustworthy, ethical, transparent, and aligned with Egypt's national values and international commitments.

As such, these Guidelines for Trustworthy and Responsible AI, as part of the full suite of AI Governance Framework in Egypt, encourages societal adoption and drives impactful innovation of AI, in line with Egypt's Vision 2030, The Egyptian Charter and the National Artificial Intelligence Strategy (2025–2030) published by the National Council for AI.

2. Methodology

The Guidelines draw inspiration from leading international frameworks and guidelines as well as different national experiences while respectful to the Egyptian context. The study phase involved different sources including:

- International Organizations and Initiatives: such as the UN GDC, UNESCO, OECD, GPAI, ITU, The Hiroshima Process… etc.
- Global AI standards: international Standards Development Organizations – ITU, ISO, IEC – and practitioner-led bodies like IEEE. Those are building a layered system of standards, spanning technical, foundational, managerial, and socio-technical domains.
- Regional and national experiences: the EU Act, The Council of Europe's Huderia, NIST, Brazil, Finland, UK, Singapore, Japan, China, India, Korea, South Africa, Kenya, Saudi Arabia and others.

The Guidelines have then gone through public consultation with relevant stakeholders and interested parties.

3. Scope & Applicability

Who will Use and to whom these Guidelines are addressed?

All AI Actors are addressed by these Guidelines. An AI Actor can be defined as "any actor involved in at least one stage of the life cycle of the AI system, and can refer to both natural and legal persons, such as researchers, programmers, IT professionals, engineers, data scientists, decision makers, end users, companies, universities and public and private entities, among others".

The Guidelines are addressing the main Actors who can be classified based on their role as:

- **Provider/Developer (PD):** those who can act as providers (whether original or middle provider) or developers of AI systems and solutions
- **Beneficiary/User (BU):** those who can be impacted as beneficiaries or direct users

The AI actors can also be classified based on their type as:

- **Government (G):** which includes governmental agencies and the public sector
- **Enterprise (E):** which includes developers and service providers
- **Community (C):** which includes an individual, group, or civic entity

In fact, any of the G, E, and C entities can play either role (PD or BU) (Figure 1). In the typical value chain scenario, E acts as a PD while both G and C act as BU. Yet in some cases, both G and C can provide and develop AI solutions and then act as PD while E could benefit from a service from another entity (e.g., business to business interaction) and then act as BU.
Throughout this document, the Guidelines will be addressed to the Actor's role whether being PD or BU. Each entity (G, E, or C) can then identify its role(s) -based on the underlying use case- as being either PD or BU and will consequently be entitled to the relevant guidelines applicable to this role(s).

Which systems are addressed by these Guidelines?

- Standalone AI systems (e.g., General Purpose AI, Generative AI,…)
- Systems that partially include AI components or modules
- Embedded AI, particularly for infrastructure monitoring and controlling (e.g., SCADA, IoT,…)
- Hardware devices with built-in AI models (e.g., smart medical devices,…)

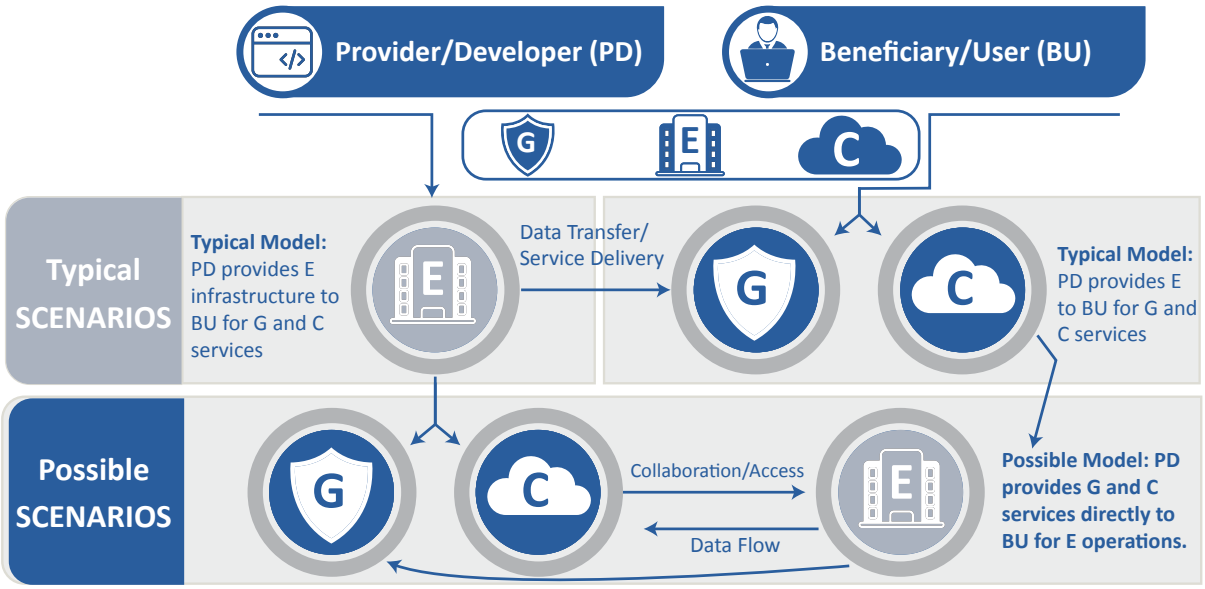


Figure 1: Organization types and roles within the AI value chain

Which parts are covered within the AI process?

The Guidelines address the whole AI system lifecycle from design to data preparation to deployment and extend to post deployment in operation phases.

Are the Guidelines generic or sector-specific?

The Guidelines are meant to be generic spanning across different sectors as a first step. Sector-specific responsible AI guidelines will be developed in due course, in collaboration with the entitled authorities within each sector.

How binding are these Guidelines?

A proportionate binding approach has been adopted[1] wherein enforcement will be proportional to the level and threshold risk. In addition, it is considered to incentivize different organizations and actors to comply with the National Guidelines for Trustworthy and Responsible AI.
It is meant to achieve the right balance between both Governance and Innovation.

1. Guide to Egypt's National AI Governance Framework

4. Assumptions

- Catastrophic failure due to cyber-attacks on an organization is addressed separately by the entitled authorities and measures. However, security measures that are AI-based or consequences of AI use are considered in these Guidelines.
- It should be noted that certain industry sectors (such as finance, healthcare, and legal) may be regulated by extra sector-specific laws or guidelines relevant to the sector itself.
- While these Guidelines are meant to be horizontal across all sectors, sector-specific guidelines may then be tailored to specific sectors and developed in collaboration with the authorities in charge and relevant stakeholders.
- While certifications (for AI management) are required at large as stated in these Guidelines, sector-specific accreditation initiatives are emerging[2] and need to be taken into consideration.
- These Guidelines will be subject to revisions based on AI technological developments and market maturity.
- These Guidelines will be followed and accompanied by a playbook that facilitates rapid uptake and implementation by the different entities.

5. Objectives and Impacts

These Guidelines are developed with the objective of:

- Helping organizations responsibly perform their role with respect to AI systems (e.g. to use, develop, monitor or provide products or services that utilize AI).
- Helping organizations to be able to ensure Trustworthy AI systems as well as assess and mitigate potential risks
- Ensuring responsible use of AI across the whole life cycle of the AI system including design, assessment and monitoring.
- Helping the community to entertain its social rights, preserve values and adopt AI systems and applications safely and in a trusted manner.
- Fostering impactful innovation that is used for good while respecting societal aspects.

2. Such as the URAC Health Care AI Accreditation, which will provide a verifiable framework for safe, ethical, and equitable AI implementation in clinical environments. In education, the Global AI Ethics in Education Charter (2025) led by UNESCO and partners, is establishing standardized codes of ethics for AI use in academia, with accreditation agencies encouraged to align local standards with this global charter

6. Concepts & Definitions

AI & AI system

AI is defined by the **ITU** as "An interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning. A computerized system that uses cognition to understand information and solve problems"[1]
ISO/IEC 2382-28 defines **AI** as "an interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning".
According to the revised definition by **OECD** in 2023, an **AI system** is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.
The **EU Act** sets further criteria for a system to be defined as **AI**, including the requirement for the system to be "designed to operate with varying levels of autonomy". This level applies to autonomous guidance and also to virtual decision processes, such as assessing the maturity of fruits, determining the quality of soil, or steering autonomous operations. To differentiate autonomous systems from automated systems, the AI system is required to possibly "exhibit adaptiveness after deployment".

AI Lifecycle

According to **the UNESCO**, an **AI system lifecycle** typically involves several phases that include: plan and design; collect and process data; build model(s) and/or adapt existing model(s) to specific tasks; test, evaluate, verify and validate; make available for use/ deploy; operate and monitor; and retire/decommission. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase. In this document, we are representing different lifecycle functions and stages under three main ones: Design & Development, Pre-deployment, and Post-Deployment.
According to **ISO/IEC 22989**, there are different roles for different players within the AI system across different lifecycle phases. These roles can include, but are not limited to one or more of the following[2]:

- **AI providers**, including AI platform providers, AI product or service providers;
- **AI producers**, including AI developers, AI designers, AI operators, AI testers and evaluators, AI deployers, AI human factor professionals, domain experts, AI impact assessors, procurers, AI governance and oversight professionals;
- **AI customers**, including AI users;
- **AI partners**, including AI system integrator and data provider;
- **AI subjects**, including data subjects and other subjects;
- Relevant authorities, including policymakers and regulators.

1. ITU (2022). Glossary – Artificial Intelligence for Natural Disaster Management. Retrieved from https://www.itu.int/dms_pub/itu-t/opb/fg/T-FG-AI4NDM-1-2022-PDF-E.pdf
2. A detailed description of these roles is provided by ISO/IEC 22989.

Organization

According to **ISO 42001**, an **Organization** is referred to a person[1] or a group of people that has its own functions with responsibilities, authorities and relationships to achieve its objectives.
In this document, an **Organization** can be classified based on the role as **Provider/ Developer (PD)** or **Beneficiary/User (BU)**. These two broad categories can represent a governmental entity (G), an enterprise (E) or a community (C) (which can be an individual or a group or a civic organization).

Management System

According to **ISO 42001**, the **Management System** is a set of interrelated or interacting elements of an organization to establish policies and objectives, as well as processes to achieve those objectives.

AI System Impact Assessment

According to **ISO 42001**, **AI System Impact Assessment** is a formal, documented process by which the impacts on individuals (and groups of individuals) and societies are identified, evaluated and addressed by an organization developing, providing or using products or services utilizing artificial intelligence.

Risk Management[2]

It should be noted that potential risks of using AI systems could cause harm to the organization itself or to other groups or the public**.** Both cases are considered in these Guidelines and need collaboration of all actors to be mitigated.
According to the **ITU report on AI governance** - 2025, **AI Risks** can be categorized into three main areas:

- **Malicious use risks**, where systems are deliberately repurposed for harmful activities, such as cyberattacks, disinformation campaigns or even the development of biological weapons;
- **Risks from malfunctions**, which arise from unforeseen technical failures, inherent biases in training data or a lack of understanding of the system's true capabilities; and
- **Systemic risks**, which encompass broader societal impacts such as market concentration, large-scale labor market disruption and the exacerbation of global inequalities.

1. The person here does not mean end user but the person who has responsibilities such as being a content creator or the owner of a company.
2. More details are included in the Risk Classification Guidelines.

Risk Sources (according to ISO/IEC 42001):

- **Complexity of environment:** When AI systems operate in complex environments, where the range of situations is broad, there can be uncertainty on the performance and therefore a source of risk (e.g. complex environment of autonomous driving).
- **Lack of transparency and Explainability:** The inability to provide appropriate information to interested parties can be a source of risk.
- **Level of automation:** Level of automation can have an impact on various areas of concerns, such as safety, fairness or security.
- **Risk sources related to machine learning:** The quality of data used for ML and the process used to collect data, can be a source of risk, as it can impact objectives such as safety and robustness (e.g. due to issues in data quality or data poisoning).
- **System hardware issues:** Risk sources related to hardware include hardware errors based on defective components or transferring trained ML models between different systems.
- **System life cycle issues:** Sources of risk can appear over the entire AI system life cycle (e.g. flaws in design, inadequate deployment, lack of maintenance, issues with decommissioning).
- **Technology readiness:** Risk sources can be related to less mature technology due to unknown factors (e.g. system limitations and boundary conditions, performance drift), but also due to the more mature technology due to technology complacency.

Frugal AI

Frugal AI is an approach for building AI systems that deliver maximum value with minimum computational, data, and energy costs. Instead of chasing ever-larger models, frugal AI focuses on **efficiency, accessibility, and sustainability** — especially for environments with limited resources.

Core Frugal AI Principles & Standards

Frugal AI is guided by principles that prioritize efficiency and sustainability throughout the entire AI lifecycle:

- **Resource Efficiency:** Minimizing computational power, energy, data, and financial capital.
- **Sustainability:** Reducing the carbon footprint through energy-efficient training and inference, as well as using sustainable data centers.
- **Accessibility & Inclusion:** Creating lightweight models that can run on low-power, edge devices (e.g., IoT, smartphones) to bridge the digital divide.
- **Impact & Scalability:** Delivering tangible, high-value outcomes using minimal resources.
- **Minimalism by Design:** Choosing the simplest, smallest model that meets performance requirements, often favoring smaller, specialized models (SLMs) over large, general-purpose ones.

Part II: Operational Guidelines

1. Overview

This Part is the operational core of the National Guidelines for Trustworthy and Responsible AI. It is built upon four foundational pillars, illustrated in Fig 2, which collectively ensure responsible and trustworthy AI adoption. These Pillars are:

- **Institutional Governance:** Organizational structures and readiness.
- **AI Lifecycle Governance:** Technical controls from design to retirement.
- **Stakeholder Engagement:** Communication and collaboration.
- **Society & Sustainability:** Ethical, cultural, and environmental alignment.

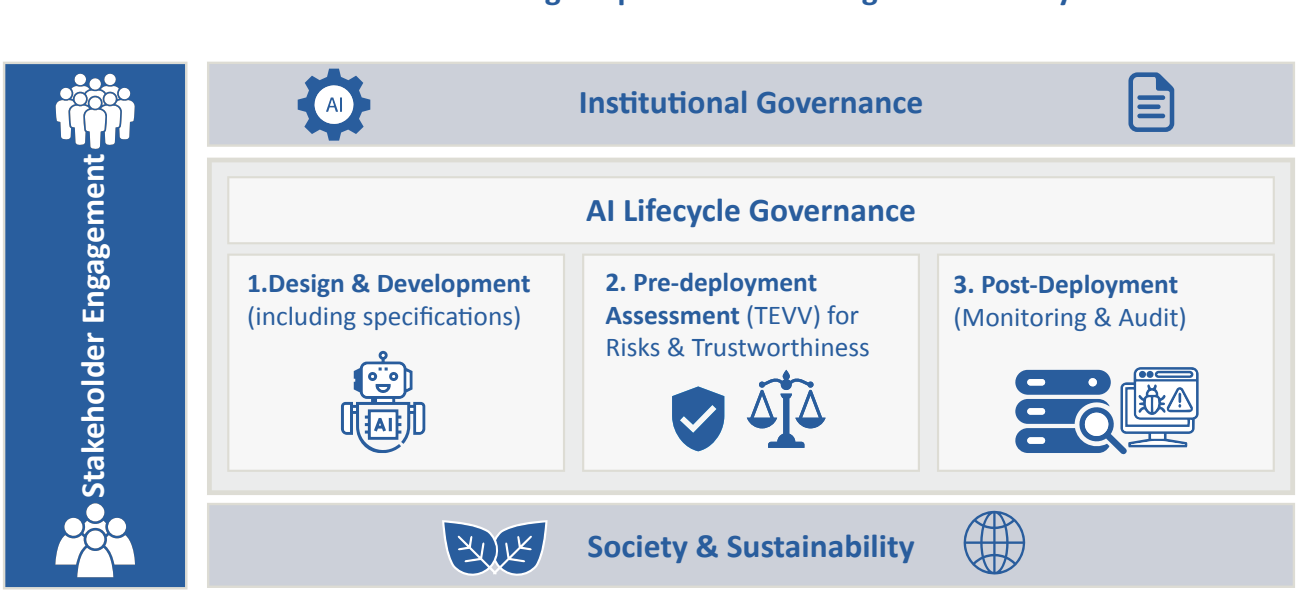


Figure 2. Egypt's Responsible AI Framework

2. Pillar I: Institutional Governance

Objective: To define how an organization is structured, managed, and internally coordinated to ensure responsible AI deployment.

This pillar serves as a prerequisite for any organization—whether a **Provider/Developer (PD)** or **Beneficiary/User (BU)**—to assess its readiness before adopting AI systems.

2.1. Policies and Procedures:

To be fully "Responsible-AI-Ready," an organization must implement procedures and policies that ensure clear governance and accountability, structured risk and technical evaluation, data and model management, human-centered oversight and transparency, strong audit, monitoring, and incident response, and a competent and well-trained workforce.

In fact, Organizations should implement the following **foundational procedures and policies** to ensure safe, ethical, and compliant use of AI:

AI Governance Policy
A high-level policy that defines:

- The organization's AI objectives and guiding principles
- Roles and responsibilities (CAIO, AI governance board, risk/compliance leads)
- AI system classification and approval pathways
- Risk thresholds and escalation procedures

AI Risk Management Procedure
The processes and documentation ensuring that all AI systems undergo **structured risk identification, assessment, mitigation, and monitoring**, including:

- Pre-deployment risk assessments
- Impact Assessment (e.g., privacy, ethics, human rights)
- Risk scoring and mitigation plans
- Continuous risk monitoring after deployment

Data Governance & Protection Policies
Procedures ensuring responsible data use, covering:

- Data quality, provenance, and lifecycle management
- Bias detection and mitigation in datasets
- Compliance with Private Data Protection Law (PDPL) and data minimization principles
- Secure data storage, access controls, and retention rules

AI Development & Engineering Standards
Technical procedures that embed responsible-by-design principles, including:

- Documentation rules (model cards, data sheets, logs)
- Testing protocols for robustness, fairness, explainability
- Adversarial testing and cybersecurity validation
- Model versioning, auditability, and reproducibility standards

Human Oversight & Operational Controls
Policies ensuring meaningful human control, such as:

- Clear oversight checkpoints (human-in-the-loop, on-the-loop)
- Operational monitoring procedures
- Decision override and escalation mechanisms
- Criteria for suspending or deactivating an AI system

AI Procurement & Vendor Governance Procedure
For acquiring third-party AI tools, organizations need:

- AI-specific procurement criteria (safety, transparency, sustainability)
- Vendor compliance documentation (TEVV results, risk reports)
- Contractual safeguards (audit rights, model update disclosures)
- Ongoing vendor monitoring requirements

Transparency & User Communication Policy
Rules ensuring clarity for users and affected individuals:

- Disclosure when interacting with AI
- Labelling of AI-generated content (watermarking for GenAI)
- User notification about data use and system limitations
- Mechanisms for explanation or appeal

Incident Response & AI Failure Management Procedure
A mandatory process for handling AI errors or harms:

- Criteria defining an "AI incident"
- Reporting, escalation, and containment steps
- Investigation procedures and corrective action plans
- Requirement to notify relevant regulators (when applicable)

Audit & Assurance Procedure
A formal process to verify responsible AI performance:

- Internal audits of high-risk systems
- Review of logs, model changes, and risk documents
- Periodic external audits when required
- Coordination with national audit entities (e.g., EGCERT, SECC-ITIDA, PDPC, ECRAI)

2.2. Workforce Readiness

Workforce readiness ensures that organizations have the skills, competencies, and institutional capacity required to oversee, manage, and implement AI systems responsibly. As AI becomes embedded across sectors, a capable socio-technical workforce is essential for maintaining accountability, ensuring compliance with national standards, and enabling safe, effective AI adoption. To achieve this, organizations must develop and sustain five core workforce readiness components:

Leadership & Accountability

Establishing clear executive responsibility—such as appointing a CAIO or governance board—to guide AI strategy, oversee risks, and ensure compliance.

Risk, Compliance & Audit Competence

Building internal capabilities to conduct TEVV assessments, maintain audit trails, and verify alignment with legal, ethical, and technical requirements.

Data & Technical Governance Skills

Ensuring personnel can manage data responsibly (quality, PDPL compliance, bias checks) and apply robust, fair, and secure AI development practices.

Human Oversight & Operational Management

Training staff to supervise AI outputs, apply escalation procedures, and maintain meaningful human control across the AI lifecycle.

Capacity Building & Ethical Alignment

Strengthening institutional knowledge through continuous training, professional development, and ethical awareness to ensure alignment with societal values and sectoral needs.

3. Pillar II: AI Lifecycle Governance

Objective: To ensure trustworthiness across the three main stages of the AI system lifecycle: **Design & Development**, **Pre-Deployment**, and **Post-Deployment**.

- **Design and Development:** this is a proactive stage that assures responsible, safe and secure AI by design from the very beginning which consequently lessens the burden and potential risks at the following stages.
- **Pre-Deployment:** represents the assessment phase to validate compliance for the whole solution before market deployment.
- **Post-Deployment:** how to keep an eye on system performance and behavior during the operational phase and make interventions as needed.

AI system implementation can start at different lifecycle stages depending on whether the solution is developed internally or procured externally (Figure 3). Internal development begins at the design phase, embedding governance, risk, and ethical controls from the outset. When adopting third-party solutions, implementation starts typically at the pre-deployment phase, emphasizing due diligence, risk assessment, validation, and contractual safeguards. Even when importing a model from a third-party, we may still need to apply some actions from the “Define and Prepare” process under the “Design and Development” stage to be able to set the requirements as well as the selection and procurement processes before moving to the pre-deployment stage. In all cases, early integration of AI governance controls is essential for safe, compliant, and effective deployment.

Across all stages of the lifecycle, the AI Principles are reflected in different fashions in one stage or more, whether partially or fully. Per instance, the Inclusiveness Principle is considered partially in both “Design” and “Embed Trustworthiness” steps under “Design and Development” stage, then assessed and verified in the “Pre-Deployment” stage, and finally monitored in real operations in the “Post-Deployment” stage.

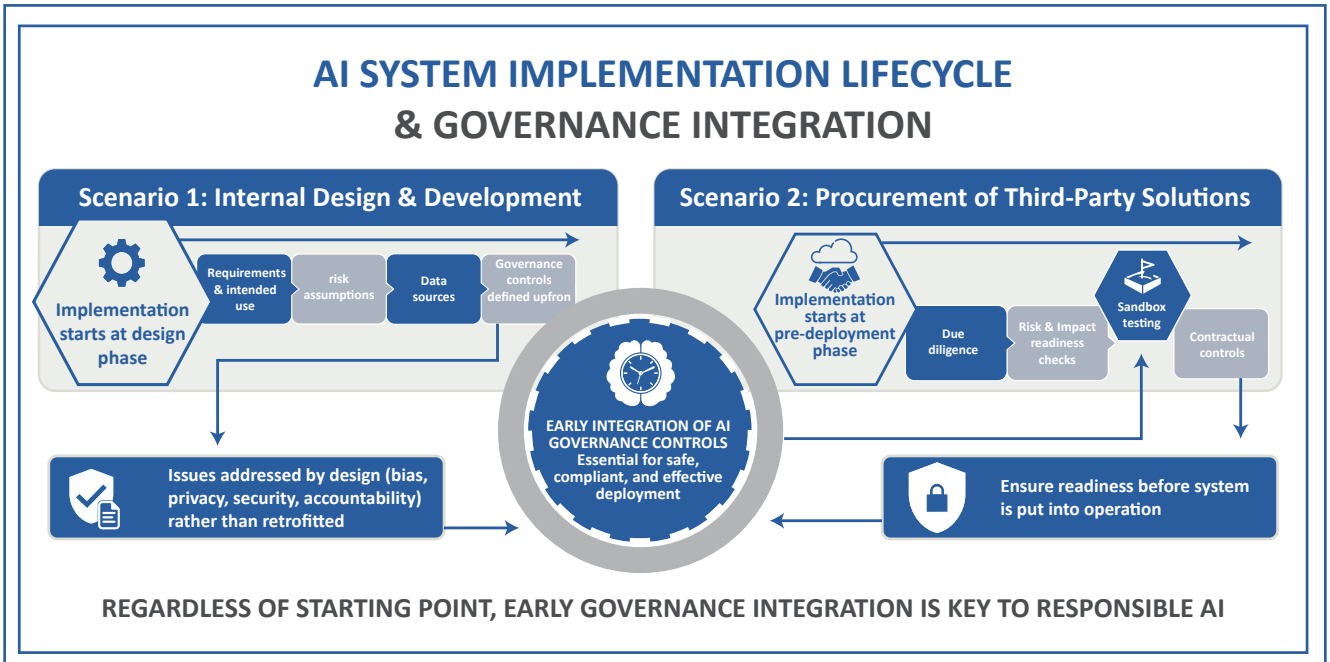


Figure 3: AI System Implementation Life Cycle

3.1. Stage 1: Design and Development

This proactive stage embeds safety and security by design (Figure 4). The design and development phase of an AI system focuses on shaping the system's architecture, functionalities and systems specifications, selecting suitable models, and ensuring that ethical, legal, and performance requirements are embedded from the start. In particular, this stage consists of three iterative cycles:

Define & Prepare

Before building, Providers/Developers and Beneficiaries/Users should collaborate to:

- **Define Purpose and Objectives:** Articulate the scope, intended uses, and end-to-end system objectives.
- **Establish Requirements:** Define product specifications and capabilities, functional needs (accuracy, latency,…), non-functional needs (training, scalability,…), and alignment with national regulations and organizational governance.
- **Identify Stakeholders:** Map all user groups, impacted parties, and vulnerable populations. Providers/Developers and Beneficiaries/Users collaborate to define roles, access levels, responsibilities, and decision rights for each group, ensuring coverage of operational, technical, regulatory, and business perspectives.
- **Assess Impact:** assess potential risks and harm scenarios for different groups and stakeholders.

Design and Build

Translating requirements into technical solutions:

- **Model and Algorithm Selection:** Select models and algorithms based on task requirements, data availability, accuracy needs, resource constraints, and environmental impact, with clear justification for choices.
- **Data Quality & Readiness:** Select relevant, representative, and unbiased data. Apply validation, preprocessing, documentation, and continuous quality checks to ensure fairness, generalizability, and robustness.
- **Frugal AI Considerations:** Favor lightweight, efficient solutions where feasible to balance performance, cost, energy consumption, and scalability. It should be noted that Frugal AI is not meant to be a mandatory choice, yet it is preferred and even recommended if the target performance and outcomes are not jeopardized.
- **Prototyping and Architecture Design:** Build prototypes to validate assumptions, system architecture, and integration pathways early, enabling rapid feedback and refinement.

Embed Trustworthiness

The following attributes must be integrated into the system architecture:

- **Fairness & Incisiveness:** Implement bias detection and mitigation to ensure equitable treatment.
- **Transparency & Explainability:** Ensure the system's capabilities and limitations are disclosed. Decisions must be explainable to users and regulators.
- **Privacy & Data Protection:** Apply Privacy-by-Design, ensuring lawful processing and protection of Personal Identifiable Information (PII).
- **Security & Robustness:** Implement safeguards against adversarial attacks and ensure fail-safe operations.
- **Accountability & Human Oversight:** Define the level of human involvement:
  - **Human-In-The-Loop (HITL):** Human retains full control; mandatory for high-risk decisions.
  - **Human-Over-The-Loop (HOTL):** Human supervises and intervenes during unexpected events.
  - **Human-Out-Of-The-Loop (HOOTL):** System is fully automated (only for low-risk scenarios).

Iterate & Ensure

- **Performance Requirements Validation:** Evaluate compliance with the product objectives, provided requirements, pre-defined acceptance metrics, and regulatory guidelines through internal auditing before moving to the Pre-deployment phase.

- **Iterative Refinement:** Use feedback, testing results, and monitoring insights to iteratively improve models, data, and controls throughout the system lifecycle.

Section 2.A.1 in Annex 1 explains detailed guidelines related to "Design and Development" and the associated assessment scheme.

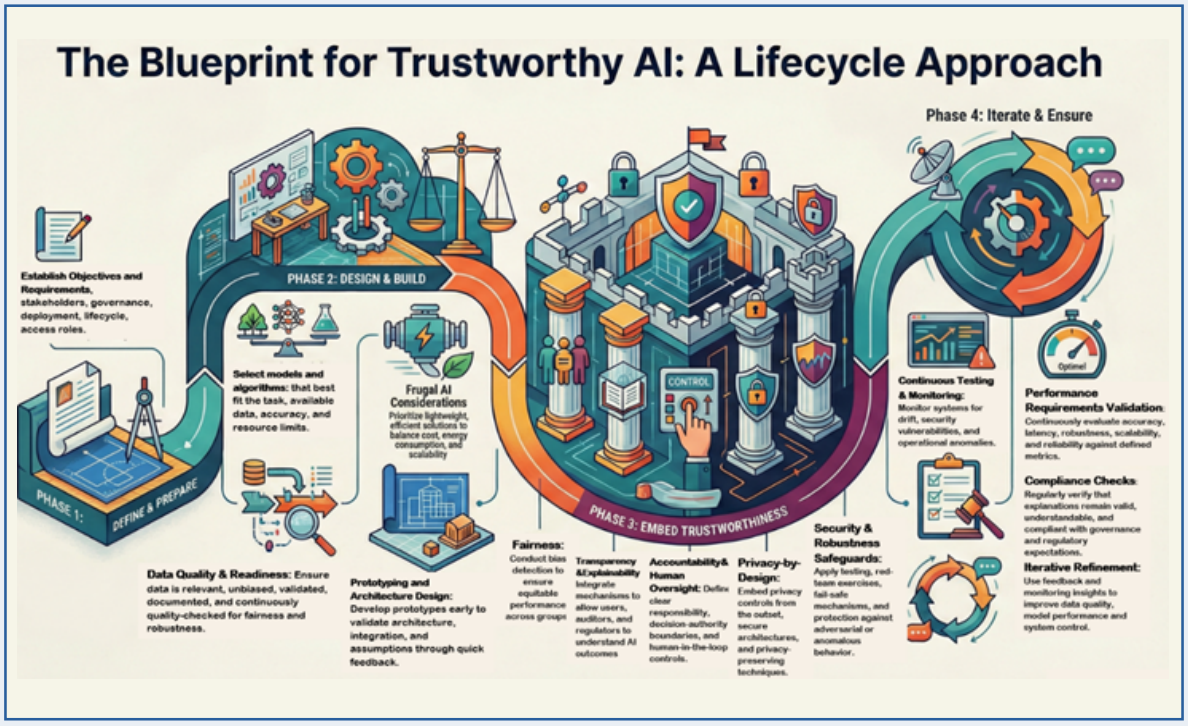


Figure 4: Design and Development Stage

3.2. Stage 2: Pre-Deployment Assessment (TEVV)

Before the AI system is released into its operational environment, a comprehensive pre-deployment assessment must be conducted using established TEVV practices[1]—Testing, Evaluation, Verification, and Validation ensuring Trustworthiness —to ensure the system is safe, reliable, and aligned with organizational and regulatory expectations (Figure 5). This assessment ensures that the system is proactively immune to potential risks and is inherently implying the trustworthy aspects. It verifies that the model performs consistently across intended use cases, edge cases, and stress scenarios, while also evaluating robustness, fairness, security, and privacy protections. It includes validating data quality, confirming that safeguards and human-oversight mechanisms function as designed, and ensuring the system meets defined performance thresholds, risk tolerances, and compliance requirements. Any identified risks must be documented, mitigated, and re-tested before deployment, ensuring the AI system demonstrates a trustworthy level of behavior and readiness for real-world operation.

- **Formal TEVV Governance & Planning:** Establish a structured TEVV plan defining scope, objectives, roles, timelines, resources, test environments, datasets, and evaluation methods, supported by a controlled regulatory sandbox where appropriate to enable supervised, pre-deployment testing under realistic conditions.
- **Standards Alignment & Regulatory Readiness:** Ensure TEVV activities align with applicable sector regulations, national laws, and recognized standards, with documentation designed for auditability, regulatory inspection, and external assurance.
- **Independent Review & Objectivity:** Integrate independent reviewers (external or functionally independent internal experts) to minimize bias and conflicts of interest, with formal documentation of findings, required remediations, and closure evidence.
- **Performance & Functional Validation:** Verify model performance against approved thresholds (accuracy, latency, robustness, scalability, uncertainty) across normal, edge, and stress conditions, ensuring safe degradation and reproducibility across environments.
- **Robustness, Security & Safety Assurance:** Validate resilience to noise, perturbation, and unexpected inputs; test resistance to adversarial and cyber threats; and confirm correct operation of fail-safe, fallback, and safe-mode mechanisms.

1. TEVV: Testing covers basic system performance; Evaluation ensures compliance with pre-defined metrics; Verification ensures that the system is built correctly and meets its specified requirements; Validation ensures meeting the needs of the users and fitting the intended purpose.

- **Fairness, Bias & Ethical Compliance:** Assess disparate impact and performance across demographic and contextual groups, document ethical and societal risks, and confirm alignment with organizational values and external ethical frameworks, including documented trade-offs.
- **Data Quality, Integrity & Privacy Controls:** Verify data completeness, accuracy, representativeness, lineage, and bias controls; confirm privacy-by-design, data minimization, retention, and lawful processing; and test with real-world and synthetic datasets to expose hidden risks.
- **Human Oversight & Safeguards Validation:** Test human-in-the-loop and human-on-the-loop mechanisms, escalation paths, monitoring alerts, and user-facing transparency and explanation features to ensure effectiveness under real operational conditions.
- **Risk Management, Traceability & Approval:** Maintain comprehensive TEVV documentation linking requirements to risks, tests, results, mitigations, and approvals; ensure no high or critical risks remain unresolved; and require formal sign-off by leadership, risk, security, and compliance owners before deployment.

Section 2.A.2 in Annex 1 provides detailed guidelines related to the “Pre-deployment” stage and its associated assessment scheme.

Figure 5: AI TEVV Process & Compliance Framework

3.3. Stage 3: Post-Deployment (Monitoring and Audit)

This stage complements the AI lifecycle after design and assessment to be able to track how the system behaves in real-world settings and being able to respond to unintended consequences or failures. In addition, AI systems that perform continuous learning change their behavior during use. Such systems require special consideration to ensure that their responsible use continues with changing behavior. This stage encompasses:

- **Real-World Performance Monitoring:** Continuous tracking of accuracy, latency, robustness, reliability, and system health in live environments, with alerts and dashboards to detect degradation, drift, or anomalies.
- **Risk, Safety & Trustworthiness Assurance:** Ongoing monitoring of fairness, bias, security threats, explainability, transparency, privacy, resilience, and effectiveness of human oversight, with automated safeguards and fallback mechanisms when risks emerge.
- **Operational & Technical Observability:** Continuous oversight of data quality, system dependencies, and data/concept drifts as well as detailed logging to support audits, investigations, and root-cause analysis.
- **Audit, Compliance & Accountability:** Regular technical, ethical, privacy, and security audits to ensure alignment with regulations, governance standards, and approved specifications, with formal documentation and stakeholder reporting.
- **Corrective Action & Continuous Improvement:** Structured feedback loops that turn monitoring results, audits, and user input into controlled updates, retraining, risk reassessment, and organizational learning across the AI lifecycle.

While the system should be able to detect any sort of concept/data drifts or anomalies when performance or outputs fall outside nominal thresholds and constraints, mitigation responses should also be applied whether by human oversight or through automated recovery responses.

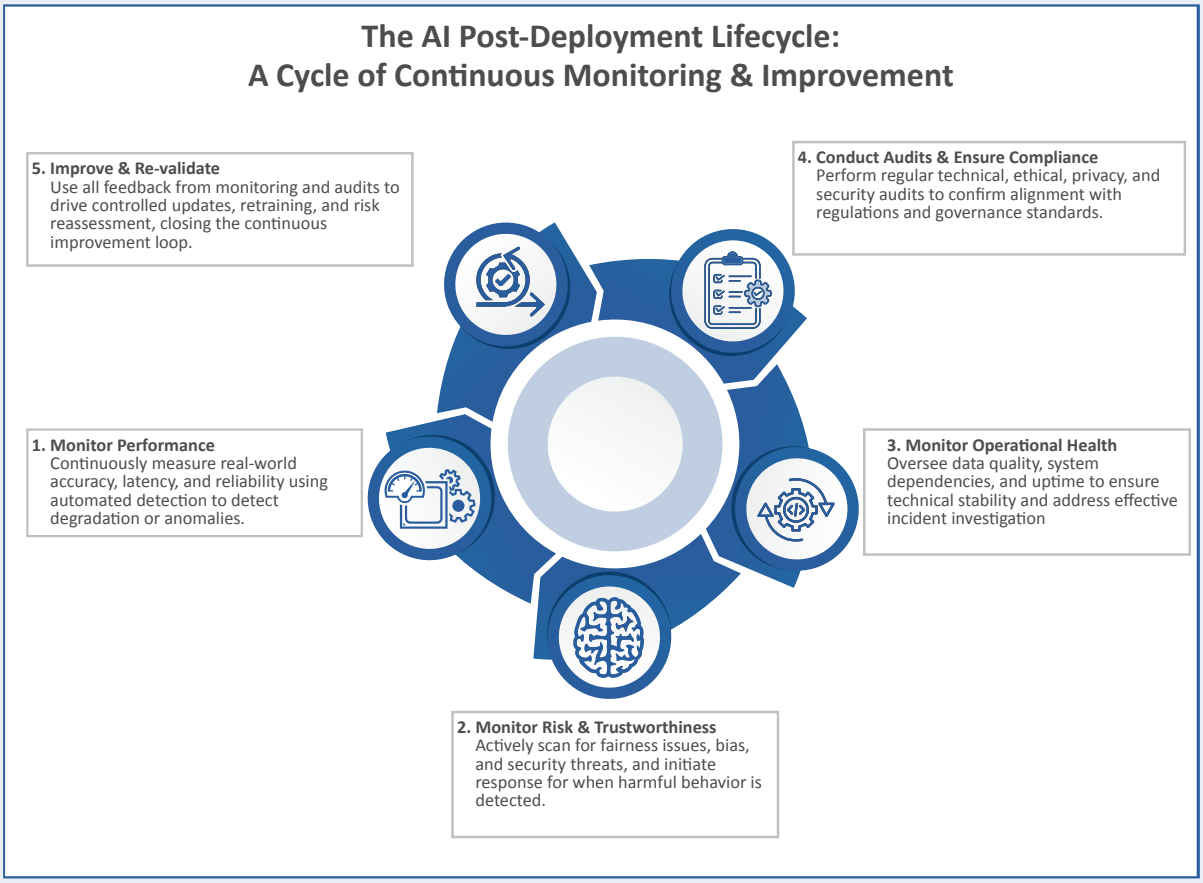


Figure 6: AI Post-Deployment Lifecycle

Section 2.A.3 in Annex 1 provides detailed guidelines related to the “Post-Deployment” stage and its associated assessment scheme.

4. Pillar III: Stakeholder Engagement

Objective: To ensure meaningful involvement of all relevant parties to foster transparency and trust. This pillar emphasizes the meaningful involvement of all relevant parties throughout the AI lifecycle to ensure effectiveness, integrity, and socially beneficial outcomes. It includes the following directions:

- **Identify, analyze, and engage stakeholders:** Map and communicate with all relevant stakeholder groups, using AI-enabled public consultation tools to make participation more inclusive and scalable. Citizens should be empowered to use complaint and appeal channels and report potential bias or discriminatory patterns to authorities or oversight bodies.
- **Foster collaboration to build trustworthy AI:** Encourage coordinated engagement among government, industry, academia, and civil society. Stakeholders should have the ability to request explanations or traceability for AI-driven decisions and to trigger audits through complaints, consultations, or collective feedback.
- **Enable a two-way communication with communities:** Provide education and awareness about AI systems while actively gathering community needs and feedback before and after deployment. Engage diverse internal and external groups to surface concerns; Understand benefits and harms, and incorporate their input into the risk management process. Establish clear complaint/appeal and consultation communication channels as well as redress mechanisms to ensure effective community engagement and feedback.
- **Ensure transparent communication on AI systems:** Clearly report system capabilities, limitations, performance metrics, safety considerations, and societal risk assessments, as well as acceptable and prohibited uses.
- **Support user reporting and responsible disclosure:** Encourage mechanisms such as issue-reporting channels, bug bounties, contests, or incentive programs to facilitate the responsible identification and disclosure of vulnerabilities or unintended system behaviors.

Section 2.B in Annex 1 provides detailed guidelines related to the “Stakeholder Engagement” Pillar and its associated assessment scheme.

5. Pillar IV: Society and Sustainability

Objective: To augment technical governance with social and environmental responsibility, this pillar is concerned with social and environmental aspects essential to augment the wholistic perspective of AI.

5.1. Societal Protection

Government entities, developers, and communities should ensure AI is adopted in a manner that respects societal values, protects individuals, and promotes responsible use. This includes:

- **Purposeful Adoption:** Use AI only when it adds clear value, rather than adopting it as a trend. Human judgment and critical thinking should remain the starting point, with AI serving to refine or enhance—not replace—human reasoning.
- **Protection of Minors:** Ensure digital and AI-enabled services are age-appropriate, include safeguards for children's data, and protect young users from harm in accordance with child protection standards.
- **Cultural and Social Alignment:** Adapt AI deployments to local societal norms, values, and expectations. Avoid importing systems designed for different cultural contexts without proper adaptation[1].
- **Transparency and User Rights:** Individuals should know when AI is being used and have access to clear explanations for consequential decisions. They must also be able to exercise privacy rights, including access, correction, and withdrawal of personal data.
- **Human-Centered Decision Making:** Maintain human judgment in all AI-assisted decisions. AI should support decisions, not autonomously replace critical human roles, especially in sensitive domains.
- **Responsible Use Boundaries:** Ensure AI is used within appropriate limits (e.g., driver-assist systems do not replace trained drivers; symptom checkers do not replace medical professionals). Communicate these boundaries clearly to users.
- **Privacy-Conscious Behavior:** Encourage prudent data sharing and informed consent, particularly regarding sensitive information such as voice, facial data, and contacts.

1. Proper adaptation includes for example customization of the training data set and relevant social aspects or limiting the use and deployment for specific cases; as AI systems and use cases may not be directly adopted as is from one country or region to another.

5.2. Environmental Protection

Once the decision is made to deploy an AI solution, government entities and enterprises should ensure that adoption is carried out in an environmentally responsible and resource-efficient manner. The use of Frugal AI is strongly recommended, whereby organizations evaluate lightweight models, energy-efficient training methods, and minimal-data techniques before selecting more complex alternatives. This approach supports sustainability while maintaining required performance standards. In practice, this includes:

- **Right-size the AI solution:** Choose the simplest model that meets the functional needs. Avoid unnecessarily large or resource-intensive models when lighter alternatives can achieve similar outcomes.
- **Adopt Frugal AI practices:** Prioritize techniques that reduce computational demand—such as model compression, pruning, quantization, and efficient architectures.
- **Use energy-efficient training methods:** Apply optimized training routines, batch tuning, scheduled training windows, and resource-efficient hardware to minimize energy consumption.
- **Minimize data requirements:** Use data-efficient approaches such as transfer learning, synthetic data where appropriate, and targeted datasets rather than excessively large corpora.
- **Evaluate environmental impact:** Require all AI proposals to include an assessment of energy use, compute requirements, and carbon footprint.
- **Select sustainable infrastructure:** Favor cloud services, data centers, and hardware accredited for energy efficiency or powered by renewable energy.
- **Monitor resource usage:** Track model runtime costs, electricity consumption, and compute utilization throughout the AI lifecycle.
- **Balance performance with sustainability:** Only select higher-complexity models when the performance gain is justified and no frugal option meets the necessary accuracy, safety, or fairness requirements.

Section 2.C. in Annex 1 provides detailed guidelines related to the “Society and Sustainability” Pillar and its associated assessment scheme.

Part III: Responsible AI Level Evaluation Methodology

The evaluation methodology establishes a unified framework for assessing the maturity and responsibility of both AI systems and the organizations that deploy them. It measures two complementary dimensions: organizational readiness (organization readiness level) and technical readiness (system-level readiness). The resulting scores provide a clear indication of the strength of responsible AI practices, the reliability of deployed AI systems, and the institutional capacity underpinning their development and use.
The methodology is composed of **two primary assessment dimensions**, each addressing a distinct dimension of responsible AI:

Enterprise Readiness Level

Enterprise readiness reflects the overall maturity of the organization deploying or operating the AI system. It focuses on **Institutional Governance**, the pillar that evaluates organizational structures, workforce capability, policies, procedures, and oversight mechanisms (Annex 1). The Institutional Governance score—out of **100 points** —determines one of the following readiness levels:

- **0–40:** Low Readiness — High institutional risk; urgent remediation required before deploying AI.
- **41–70:** Moderate Readiness — Governance foundations exist but remain incomplete; improvements needed.
- **71–100:** High Readiness — Strong governance capabilities; organization is structurally equipped to deploy AI responsibly.

This ensures that even a well-governed AI system is **not approved for deployment** unless the enterprise operating it demonstrates adequate readiness.

Responsible AI System Readiness Level

The AI system readiness score reflects the AI system's alignment with the four pillars of the national governance framework. It provides a **holistic maturity indicator** that moves beyond compliance to highlight the system's operational trustworthiness.

The total combined score is **380 points (**scoring details is presented in Annex 1**)**, distributed across the pillars as follows:

#	Pillar	Max
2	AI Life Cycle Governance	274
3	Stakeholders Engagement	40
4	Societal & Sustainability	66
Overall Score		380

Readiness Levels (based on total score)

- **0–152:** Low Readiness — High risk; major shortcomings in governance and system trustworthiness.
- **153–266:** Moderate Readiness — Partial alignment; targeted improvements required prior to scaled deployment.
- **267–380:** High Readiness — Strong responsible AI practices embedded; system is reliable, trustworthy, and well-governed.

This scoring model ensures that readiness is measured not only by compliance, but by the **maturity of responsible AI practices across the entire ecosystem of the system.**

Summary of the Responsible AI Evaluation Model

Parameter	Purpose	Output
Enterprise Readiness	Assesses organizational governance maturity	Score 0–100 + Institutional Readiness Level
System Readiness Score	Evaluates trustworthiness across all framework pillars	Score 0–380 + AI System Readiness Level

Part IV: Closing Note

These Guidelines establish AI Governance as an ongoing public responsibility—one that extends far beyond a single technical assessment or compliance checkpoint. Effective governance requires continuous vigilance, regular review, and active collaboration across institutions to ensure that AI systems remain safe, lawful, and aligned with national values as they evolve.

Consistent application of these principles is essential to sustaining public trust, strengthening institutional accountability, and ensuring that AI is used in ways that protect citizens' rights and advance the public good. As technologies change and new challenges emerge, government entities must remain adaptable, transparent, and ethically grounded.

Ultimately, Responsible AI Governance is not merely a regulatory commitment—it is a national stewardship obligation. By upholding these standards, Egypt ensures that AI innovations contribute to societal well-being, economic competitiveness, and a future in which technology serves people with fairness, safety, and integrity.

Part IV: Annexes

Annex 1: Responsible AI Evaluation Tool

1. Enterprise Readiness Level

EG Responsible AI Institutional Governance – Self-Assessment Scoring Tool
For Government Entities
This Institutional Governance Checklist Should Be Completed BEFORE Deploying, Procuring, or Expanding Any AI System.
Scoring Scale

Score	Meaning	Description
0 = Not Implemented	No evidence, not started	Policy/function is missing or undocumented
1 = Partially Implemented	Some progress	Exists in draft or pilot form; not institutionalized or consistently applied
2 = Fully Implemented	Complete & operational	Clear documentation, assigned responsibilities, fully functioning and monitored

Maximum Score: 100 points
Readiness Levels:

- **0–40 = Low Readiness** (Significant gaps, high risk)
- **41–70 = Moderate Readiness** (Some structures in place, needs improvement)
- **71–100 = High Readiness** (Strong governance foundation, compliant with guidelines)

SECTION 1 — Leadership, Roles & Accountability (20 points)

Assessment Item	Score (0–2)
1. A CAIO or equivalent senior leader is formally appointed.	☐ 0 ☐ 1 ☐ 2
2. An AI Governance Committee exists with documented mandate.	☐ 0 ☐ 1 ☐ 2
3. Roles are assigned for system ownership, risk, data, and oversight.	☐ 0 ☐ 1 ☐ 2
4. AI decision-making authority is documented and approved.	☐ 0 ☐ 1 ☐ 2
5. Governance structure aligns with national bodies (NCAI, regulators).	☐ 0 ☐ 1 ☐ 2
6. Internal reporting lines for AI accountability are clear.	☐ 0 ☐ 1 ☐ 2
7. Leadership reviews AI risks regularly.	☐ 0 ☐ 1 ☐ 2
8. Executive oversight includes ethics, fairness, and safety.	☐ 0 ☐ 1 ☐ 2
9. AI responsibilities are integrated into job descriptions.	☐ 0 ☐ 1 ☐ 2
10. Succession planning exists for key AI governance roles.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 20): _____

SECTION 2 — Policies & Procedures (20 points)

Required Policy / Procedure	Score (0–2)
11. AI Governance Policy adopted and communicated.	☐ 0 ☐ 1 ☐ 2
12. AI Risk Management Procedure (includes TEVV).	☐ 0 ☐ 1 ☐ 2
13. Data Governance & PDPL Compliance Policy.	☐ 0 ☐ 1 ☐ 2
14. Human Oversight & Control Policy.	☐ 0 ☐ 1 ☐ 2
15. AI Procurement & Vendor Governance Policy.	☐ 0 ☐ 1 ☐ 2
16. Transparency & User Communication Policy.	☐ 0 ☐ 1 ☐ 2
17. AI Incident Response Procedure.	☐ 0 ☐ 1 ☐ 2
18. Monitoring, Audit & Continuous Improvement Procedure.	☐ 0 ☐ 1 ☐ 2
19. Workforce Competency & Training Policy.	☐ 0 ☐ 1 ☐ 2
20. Ethical & Societal Impact Assessment Procedure.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 20): _____

SECTION 3 — AI Inventory & Risk Classification (10 points)

Assessment Item	Score
21. Complete inventory of all AI systems (existing and planned).	☐ 0 ☐ 1 ☐ 2
22. AI systems classified by risk tier (Tier 1–4).	☐ 0 ☐ 1 ☐ 2
23. GPAI / LLM-based systems are flagged for special governance.	☐ 0 ☐ 1 ☐ 2
24. High-risk systems require dual-check TEVV.	☐ 0 ☐ 1 ☐ 2
25. Each AI system has an assigned owner.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 10): _____

SECTION 4 — Workforce Readiness & Capacity (10 points)

Assessment Item	Score
26. Governance leaders trained (CAIO, committee).	☐ 0 ☐ 1 ☐ 2
27. Operators trained to challenge and override AI outputs.	☐ 0 ☐ 1 ☐ 2
28. Procurement teams trained on AI requirements.	☐ 0 ☐ 1 ☐ 2
29. Legal & compliance staff trained on AI ethics and PDPL.	☐ 0 ☐ 1 ☐ 2
30. Technical staff trained on TEVV, security, fairness, drift.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 10): _____

SECTION 5 — Data Governance & Protection (20 points)

Assessment Item	Score
31. Data provenance documented for each system.	☐ 0 ☐ 1 ☐ 2
32. Data lineage traceability maintained for auditability and reproducibility.	☐ 0 ☐ 1 ☐ 2
33. Data quality validated (accuracy, completeness, relevance, consistency).	☐ 0 ☐ 1 ☐ 2
34. Ongoing data quality monitoring implemented with defined KPIs.	☐ 0 ☐ 1 ☐ 2
35. Bias detection and mitigation applied to datasets.	☐ 0 ☐ 1 ☐ 2

36. PDPL compliance ensured (minimization, consent, retention).	☐ 0 ☐ 1 ☐ 2
37. Privacy-by-design and privacy-by-default applied across the data lifecycle.	☐ 0 ☐ 1 ☐ 2
38. Secure data handling practices applied (encryption, access control, logging).	☐ 0 ☐ 1 ☐ 2
39. Regular data governance audits and risk reviews conducted.	☐ 0 ☐ 1 ☐ 2
40. Data access and usage governed by defined roles, approvals, and accountability mechanisms.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 20): _____

SECTION 6 — Human Oversight & Accountability (10 points)

Assessment Item	Score
41. Oversight model defined (HITL / HOTL) for each system.	☐ 0 ☐ 1 ☐ 2
42. Manual override mechanisms exist and are tested.	☐ 0 ☐ 1 ☐ 2
43. Escalation pathways documented.	☐ 0 ☐ 1 ☐ 2
44. Operators can interpret and contest AI outputs.	☐ 0 ☐ 1 ☐ 2
45. Oversight decisions and interventions are logged.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 10): _____

SECTION 7 — Monitoring, Audit & Continuous Improvement (10 points)

Assessment Item	Score
46. Monitoring dashboards exist for performance and drift.	☐ 0 ☐ 1 ☐ 2
47. Regular internal audits conducted for high-risk systems.	☐ 0 ☐ 1 ☐ 2
48. External audits conducted as required.	☐ 0 ☐ 1 ☐ 2
49. Audit findings lead to corrective actions.	☐ 0 ☐ 1 ☐ 2
50. Lessons learned feed back into policy updates.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 10): _____

TOTAL SCORE (out of 100)	
Leadership & Accountability:	/ 20
Policies & Procedures:	/ 20
AI Inventory & Classification:	/ 10
Workforce Readiness:	/ 10
Data Governance:	/ 20
Human Oversight:	/ 10
Monitoring & Audit:	/ 10

- **0–40:** Low Readiness — high risk, immediate remediation required
- **41–70:** Moderate Readiness — governance improving but incomplete
- **71–100:** High Readiness — strong institutional governance foundation

2. Responsible AI System Readiness

A. AI systems Life Cycle Governance

The AI Systems Life Cycle Governance Checklist Should Be Completed before the implementation of any AI system.

Scoring Scale

Score	Meaning	Description
0 = Not Implemented	No evidence, not started	Policy/function/item is missing or undocumented
1 = Partially Implemented	Some progress	Exists in draft or pilot form; not institutionalized or consistently applied
2 = Fully Implemented	Complete & operational	Clear documentation, assigned responsibilities, fully functioning and monitored

Maximum Score: 274 points

Readiness Levels:

- **0–109:** Low Readiness — High risk; major shortcomings in governance and system trustworthiness.
- **110–191:** Moderate Readiness — Partial alignment; targeted improvements required prior to scaled deployment.
- **192–274:** High Readiness — Strong responsible AI practices embedded; system is reliable, trustworthy, and well-governed.

A.1 DESIGN & DEVELOPMENT

A.1.1 DEFINE AND PREPARE

Assessment Item	Score
A. Objectives, Functionalities & System Requirements	
1. Clearly document the AI system's intended purpose, scope, and expected outcomes, including intended and non-intended uses.	☐ 0 ☐ 1 ☐ 2
2. Define end-to-end functional capabilities and non-functional requirements (performance, reliability, security, scalability, resilience).	☐ 0 ☐ 1 ☐ 2
3. Specify deployment context and constraints (cloud, edge, on-prem), integration dependencies, and interoperability requirements.	☐ 0 ☐ 1 ☐ 2
4. Define lifecycle management requirements, including monitoring, retraining, versioning, rollback, and decommissioning.	☐ 0 ☐ 1 ☐ 2
5. Model selection must align with the system's risk level, prioritizing interpretability, robustness, and auditability for higher-risk applications. Choices must be feasible within data, computational, domain, and environmental constraints, with explicit comparison between frugal and larger models and clear justification for the final selection.	☐ 0 ☐ 1 ☐ 2
6. Identify target user groups, usage contexts, and decision criticality, with specifications jointly validated by Provider/ Developer and Business Units (BU).	☐ 0 ☐ 1 ☐ 2
7. Ensure alignment with applicable national regulations, sectoral rules, and internal AI governance policies before design approval.	☐ 0 ☐ 1 ☐ 2
B. Data Quality & Readiness	
8. Identify and document data sources, ownership, provenance, and legal basis for use.	☐ 0 ☐ 1 ☐ 2
9. Ensure datasets are relevant, representative, and suitable for the intended purpose and risk level of the system.	☐ 0 ☐ 1 ☐ 2

10. Apply standardized data validation, preprocessing, labeling, and documentation practices.	☐ 0 ☐ 1 ☐ 2
11. Conduct bias, imbalance, and coverage assessments to reduce unfair or discriminatory outcomes.	☐ 0 ☐ 1 ☐ 2
12. Define continuous data quality monitoring and update mechanisms throughout the lifecycle.	☐ 0 ☐ 1 ☐ 2
C. User & Stakeholder Groups	
13. Clearly identify and document all relevant user and stakeholder groups, including primary users, secondary users, operators, administrators, decision-makers, beneficiaries, and indirectly affected or impacted groups.	☐ 0 ☐ 1 ☐ 2
14. Document the intended purpose of the AI system, its anticipated benefits, applicable laws and norms, and the operational, organizational, and social environments in which it will be used.	☐ 0 ☐ 1 ☐ 2
15. Define usage contexts, decision criticality, and human–AI interaction modes for each group, including whether the system supports, informs, or automates decisions.	☐ 0 ☐ 1 ☐ 2
16. Specify roles, responsibilities, access rights, and accountability for each group, ensuring clear ownership across Provider/Developer, Business Units (BU), and operational teams.	☐ 0 ☐ 1 ☐ 2
17. Assess potential impacts, risks, and harm scenarios for different stakeholder groups, with particular attention to vulnerable or high-impact populations.	☐ 0 ☐ 1 ☐ 2
18. Define user competence, training, and awareness requirements to ensure safe, appropriate, and effective system use.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 36): _____

A.1.2 DESIGN & BUILD

Assessment Item	Score
A. Model & Algorithm Selection	
1. Require a documented and reviewable justification for all model and algorithm selections, including task suitability, performance expectations, data characteristics, risk level, and operational constraints.	☐ 0 ☐ 1 ☐ 2
2. Mandate explicit comparison between lightweight/frugal models and larger or more resource-intensive alternatives. Selection of non-frugal models must demonstrate clear, measurable benefits that outweigh cost, energy, scalability, and sustainability impacts.	☐ 0 ☐ 1 ☐ 2
3. Enforce prioritization of interpretable and auditable models for high-risk or high-impact use cases. Use of opaque or "black-box" models require documented risk mitigation measures and formal governance approval.	☐ 0 ☐ 1 ☐ 2
4. Require assessment of computational cost, energy consumption, and infrastructure impact as part of model selection decisions.	☐ 0 ☐ 1 ☐ 2
B. Data Quality & Readiness	
5. Ensure that all datasets used for training, validation, and testing have documented provenance, ownership, licensing status, and transformation history, enabling full traceability and auditability.	☐ 0 ☐ 1 ☐ 2
6. Data Quality Assurance, mandate assessment of training data against defined quality dimensions—accuracy, completeness and representativeness, consistency, and timeliness—commensurate with system risk and decision criticality.	☐ 0 ☐ 1 ☐ 2
7. Require controls to detect and mitigate bias, imbalance, and integrity risks in data, including systematic review of sources and sampling strategies.	☐ 0 ☐ 1 ☐ 2
8. Enforce technical and procedural checks to identify potential data poisoning, malicious manipulation, or unintended bias injection prior to model training.	☐ 0 ☐ 1 ☐ 2

C. General Design & Build Controls	
9. Ensure that all design and build activities strictly implement the approved objectives, system requirements, risk classification, and governance decisions defined in earlier phases.	☐ 0 ☐ 1 ☐ 2
10. Apply trustworthiness-by-design principles, embedding accountability, security, robustness, privacy, and transparency into technical choices from the outset.	☐ 0 ☐ 1 ☐ 2
11. Prohibit ad-hoc or undocumented deviations from approved designs without formal governance review and approval.	☐ 0 ☐ 1 ☐ 2
D. Prototyping and Validation	
12. Require early prototyping to validate assumptions related to model behavior, system architecture, data suitability, and integration pathways.	☐ 0 ☐ 1 ☐ 2
13. Use prototypes to identify technical, ethical, and operational risks before full-scale development.	☐ 0 ☐ 1 ☐ 2
14. Ensure that lessons learned from prototyping are formally documented and incorporated into design decisions.	☐ 0 ☐ 1 ☐ 2
Oversight, Review, and Compliance	
15. Subject design and build outputs to formal governance review before deployment, including model choice, data readiness, and risk mitigation measures.	☐ 0 ☐ 1 ☐ 2
16. Ensure alignment with applicable regulations, sectoral standards, and internal AI governance policies prior to release.	☐ 0 ☐ 1 ☐ 2
17. Maintain auditable records of all design decisions, approvals, exceptions, and risk acceptances.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 34): _____

A.1.3 EMBED TRUSTWORTHINESS

Assessment Item	Score
Fairness and Inclusiveness	
1. Ensure AI systems are designed, trained, and evaluated to treat individuals and groups equitably, without unjustified bias or discrimination.	☐ 0 ☐ 1 ☐ 2
2. Identify, assess, and mitigate potential biases related to gender, age, ethnicity, disability, socioeconomic status, geography, or other protected or contextually relevant characteristics across the full lifecycle.	☐ 0 ☐ 1 ☐ 2
3. Apply fairness-aware data practices, including representative data collection, bias and imbalance testing, and documented mitigation strategies.	☐ 0 ☐ 1 ☐ 2
4. Prevent unequal or discriminatory treatment by defining fairness metrics, thresholds, and acceptance criteria appropriate to the system's context and risk level.	☐ 0 ☐ 1 ☐ 2
5. Where applicable, involve diverse communities and domain experts in requirement definition, data selection, testing, and evaluation to reflect real-world diversity.	☐ 0 ☐ 1 ☐ 2
6. Ensure inclusiveness by considering accessibility needs and usability requirements for persons with disabilities or special needs, in line with recognized accessibility standards.	☐ 0 ☐ 1 ☐ 2
7. Document fairness and inclusiveness decisions, assessments, and trade-offs, and subject them to governance review and approval before deployment and during significant system updates.	☐ 0 ☐ 1 ☐ 2

Transparency and Explainability	
8. Assign clear responsibilities, maintain traceability, and enable external auditability	☐ 0 ☐ 1 ☐ 2
9. Implement mechanisms such as clear labeling, disclosures, or disclaimers to inform users, where appropriate, that they are interacting with an AI system, AI-generated content, or AI-supported decision-making.	☐ 0 ☐ 1 ☐ 2
10. Provide clear, accessible, and non-technical information describing the AI system's purpose, capabilities, limitations, expected performance, and known risks.	☐ 0 ☐ 1 ☐ 2
11. Clearly communicate the role of automation versus human involvement, including whether the system informs, supports, or autonomously executes decisions.	☐ 0 ☐ 1 ☐ 2
12. Select model architectures and Explainability techniques proportional to the system's risk category, ensuring higher-risk systems support stronger interpretability and traceability.	☐ 0 ☐ 1 ☐ 2
13. Design explanations that are appropriate for different audiences (e.g., end users, operators, regulators), avoiding misleading simplifications while remaining understandable.	☐ 0 ☐ 1 ☐ 2
14. Ensure that transparency and Explainability requirements are jointly validated by Provider/Developer and Business Units (BU) during design approval.	☐ 0 ☐ 1 ☐ 2
15. Document transparency measures, Explainability methods, and related limitations as part of the system's technical and governance documentation, and review them during audits and major system updates.	☐ 0 ☐ 1 ☐ 2
16. Ensure that transparency and Explainability requirements are jointly validated by Provider/Developer and Business Units (BU) during design approval.	☐ 0 ☐ 1 ☐ 2
17. Document transparency measures, Explainability methods, and related limitations as part of the system's technical and governance documentation, and review them during audits and major system updates.	☐ 0 ☐ 1 ☐ 2

Accountability and Human Oversight	
18. Define and document clear roles, responsibilities, and decision ownership across the AI lifecycle, including design, deployment, operation, monitoring, and incident response. Accountability must remain with designated human roles even when decisions are supported or automated by AI systems.	☐ 0 ☐ 1 ☐ 2
19. Establish a risk-based human oversight model for each AI use case, ensuring that the level of human involvement is proportionate to the system's potential harm, decision criticality, and likelihood of error.	☐ 0 ☐ 1 ☐ 2
20. Require human-in-the-loop controls (HITL) for high-risk or high-impact decisions, where AI systems provide recommendations or inputs only. Final decisions must require explicit human review and approval before execution.	☐ 0 ☐ 1 ☐ 2
21. Apply human-over-the-loop oversight (HOTH) for systems operating with partial autonomy, ensuring continuous human monitoring, clear alerting mechanisms, and the ability for authorized personnel to intervene, override, or suspend system operation in case of anomalies, failures, or unexpected outcomes.	☐ 0 ☐ 1 ☐ 2
22. Permit full automation only for low-risk (HOOTL) use cases where potential harm is minimal, well-understood, and reversible, and where adequate safeguards, monitoring, and fallback mechanisms are in place.	☐ 0 ☐ 1 ☐ 2
23. Ensure inclusiveness by considering accessibility needs and usability requirements for persons with disabilities or special needs, in line with recognized accessibility standards.	☐ 0 ☐ 1 ☐ 2
24. Determine the appropriate oversight mode (HITL, HOTL, or HOOTL) based on documented harm and risk assessments, considering severity, probability of occurrence, reversibility of impact, and affected stakeholders.	☐ 0 ☐ 1 ☐ 2
25. Define clear escalation paths, intervention thresholds, and operational procedures to ensure timely human action when system behavior deviates from expected or acceptable limits.	☐ 0 ☐ 1 ☐ 2
26. Maintain audit logs and documentation of AI outputs, human decisions, overrides, and interventions to support transparency, accountability, and post-incident analysis.	☐ 0 ☐ 1 ☐ 2
27. Regularly review accountability assignments and oversight arrangements to reflect system updates, changes in risk profile, regulatory developments, and operational experience.	☐ 0 ☐ 1 ☐ 2

Privacy	
28. Embed privacy considerations into the design, development, and operation of AI systems from the outset, ensuring that personal data protection is the default setting throughout the lifecycle.	☐ 0 ☐ 1 ☐ 2
29. Protect privacy for both directly processed personal data and indirectly inferred personal attributes, including sensitive or derived information that may impact individuals.	☐ 0 ☐ 1 ☐ 2
30. Clearly define and document the roles of personal data controllers and processors in accordance with applicable data protection laws, ensuring accountability for lawful processing, safeguards, and compliance obligations.	☐ 0 ☐ 1 ☐ 2
31. Ensure that all processing of personal data has a valid legal basis, is limited to clearly defined purposes, and is not reused or repurposed without appropriate legal justification.	☐ 0 ☐ 1 ☐ 2
32. Limit the collection, use, and retention of personal data to what is strictly necessary for the intended purpose, risk level, and system functionality.	☐ 0 ☐ 1 ☐ 2
33. Implement appropriate technical and organizational measures, including access controls, encryption, and segregation of duties, to prevent unauthorized access, leakage, or misuse of personal data.	☐ 0 ☐ 1 ☐ 2
34. Transparency toward data subjects regarding how their data is processed and support the exercise of data protection rights (e.g., access, correction, deletion), in line with applicable regulations.	☐ 0 ☐ 1 ☐ 2
35. Define clear data retention periods and secure deletion mechanisms consistent with legal, regulatory, and business requirements.	☐ 0 ☐ 1 ☐ 2
36. Ensure that third-party processors and cross-border data transfers comply with applicable privacy laws, contractual safeguards, and security standards.	☐ 0 ☐ 1 ☐ 2
37. Regularly review privacy risks, controls, and compliance measures, especially following system updates, changes in data sources, or evolving regulatory requirements.	☐ 0 ☐ 1 ☐ 2

Security and Robustness	
38. Security by Design: Integrate security controls throughout the AI system lifecycle, ensuring that adequate technical and organizational safeguards are in place to protect systems, models, data, and interfaces against unauthorized access, misuse, and cyber threats.	☐ 0 ☐ 1 ☐ 2
39. Robustness by Design: Design AI systems to operate reliably under expected and unexpected conditions, mitigating both intended and unintended failures, including data errors, model degradation, and environmental changes.	☐ 0 ☐ 1 ☐ 2
40. Identify, adopt, and comply with applicable sector-specific safety standards, certifications, and assurance requirements (e.g., Justice, medical AI, financial systems), where relevant.	☐ 0 ☐ 1 ☐ 2
41. Protect models and data against poisoning, evasion, adversarial inputs, model extraction, and unauthorized modification through secure pipelines, validation checks, and controlled access.	☐ 0 ☐ 1 ☐ 2
42. Define and validate contingency, fail-safe, and fallback plans during the design phase, including manual override or safe-mode operation, to ensure continuity and safety in the event of system malfunction or security incidents.	☐ 0 ☐ 1 ☐ 2
43. Implement architectural and operational safeguards to prevent cascading failures, unsafe system interactions, or uncontrolled responses, particularly in safety-critical or high-impact applications.	☐ 0 ☐ 1 ☐ 2
44. Establish continuous monitoring, alerting, and incident response mechanisms to detect anomalies, performance degradation, or security breaches in a timely manner.	☐ 0 ☐ 1 ☐ 2
45. Ensure the system can adapt to changes over time, such as concept drift or evolving data distributions, through controlled updates, retraining, validation, and performance monitoring without unacceptable degradation.	☐ 0 ☐ 1 ☐ 2
46. Conduct periodic robustness, stress, and security testing, including penetration testing and scenario-based evaluations, and update controls in response to new threats or operational insights.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 92): _____

A.1.4 ITERATE & ENSURE

Assessment Item	Score
Performance Requirements	
1. Define measurable performance metrics, including accuracy, latency, robustness, and scalability, aligned with the system's intended purpose, risk classification, and operational context.	☐ 0 ☐ 1 ☐ 2
2. Require validation of performance across expected, boundary, and stress scenarios before deployment and after each significant update.	☐ 0 ☐ 1 ☐ 2
3. Establish minimum acceptance thresholds and escalation procedures when performance degradation is detected.	☐ 0 ☐ 1 ☐ 2
Documentation & Governance	
4. Maintain comprehensive and up-to-date documentation covering system design, data sources, model versions, evaluation results, changes, and known limitations.	☐ 0 ☐ 1 ☐ 2
5. Ensure continuous alignment with applicable national regulations, sectoral requirements, and internal AI governance policies.	☐ 0 ☐ 1 ☐ 2
6. Preserve clear, auditable records of decisions, approvals, test results, incidents, and corrective actions throughout the system lifecycle.	☐ 0 ☐ 1 ☐ 2
Continuous Testing and Improvement	
7. Implement continuous testing processes, including functional, robustness, and stress testing, proportionate to the system's risk level.	☐ 0 ☐ 1 ☐ 2
8. Conduct periodic red-team and adversarial evaluations to identify vulnerabilities, misuse risks, and failure modes.	☐ 0 ☐ 1 ☐ 2
9. Use test results to drive controlled, documented iterative refinement, ensuring improvements do not introduce new risks or regressions.	☐ 0 ☐ 1 ☐ 2

Explainability and Transparency	
10. Ensure that the system remains sufficiently explainable for its risk level, user groups, and decision criticality.	☐ 0 ☐ 1 ☐ 2
11. Maintain mechanisms to provide meaningful explanations of system outputs to users, operators, auditors, and regulators, as applicable.	☐ 0 ☐ 1 ☐ 2
12. Reassess Explainability requirements following model updates, retraining, or changes in usage context.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 24): _____

TOTAL SCORE (out of 186)	
Define and Prepare:	**/ 36**
Design & Build:	**/ 34**
Embed Trustworthiness:	**/ 92**
Iterate & Ensure:	**/ 24**

B. Pre-Deployment Assessment (TEVV)

Assessment Item	Score
TEVV Planning & Governance Structure	
1. **Establish/Outsource AI Regulatory Sandbox** to provide a control and supervised testing environment **where AI systems can be evaluated under real or simulated conditions before final approval and deployment under lightweight regulations.**	☐ 0 ☐ 1 ☐ 2
2. **A formal TEVV plan must be established**: Defines scope, objectives, responsibilities, timelines, and required resources. Specifies testing environments, datasets, tools, and evaluation methodologies.	☐ 0 ☐ 1 ☐ 2
3. **TEVV must follow recognized standards and regulatory expectations**: Align sector regulations, and local legal requirements. Include documentation to support auditability and regulatory inspections.	☐ 0 ☐ 1 ☐ 2

4. **Independent review mechanisms must be incorporated**: Use external or uninvolved internal experts to reduce internal bias and conflicts of interest. Document independent assessments and required remediations.	☐ 0 ☐ 1 ☐ 2
Functional & Performance Validation	
5. **Validate model performance against defined thresholds:** Accuracy, latency, throughput, robustness, scalability, and uncertainty outputs. Confirm performance under realistic operational conditions.	☐ 0 ☐ 1 ☐ 2
6. **Test behavior across edge cases and stress scenarios**: Overload conditions, rare events, degraded inputs, low-quality data, and high-variance contexts. Ensure the system degrades safely when performance drops.	☐ 0 ☐ 1 ☐ 2
7. **Verify consistency of results across environments**: Ensure reproducibility across development, test, staging, and production-like environments. Document configuration differences and their impact.	☐ 0 ☐ 1 ☐ 2
Robustness, Security, and Safety Assurance	
8. **Conduct robustness testing:** Assess stability against perturbations, noise, incomplete data, or unexpected inputs.	☐ 0 ☐ 1 ☐ 2
9. **Conduct adversarial and security testing:** Resistance to adversarial attacks (evasion, poisoning, extraction). Impact of cyber threats and malicious manipulation.	☐ 0 ☐ 1 ☐ 2
10. Validate safety and fail-safe mechanisms: **Ensure fallback behaviors, safe-mode operations, and risk-triggered shutdowns operate correctly. Validate that safety thresholds produce predictable, safe responses.**	☐ 0 ☐ 1 ☐ 2

Fairness, Bias, and Ethical Compliance	
11. **Evaluate fairness across demographic and contextual variations:** Measure disparate impact, outcome disparities, and performance differences across groups. Assess fairness in both model outputs and downstream decision processes.	☐ 0 ☐ 1 ☐ 2
12. Document ethical and societal impact considerations: Assess potential harms to individuals, communities, and vulnerable groups. Validate mitigation strategies for unacceptable or disproportionate risks.	☐ 0 ☐ 1 ☐ 2
13. Confirm compliance with ethical guidelines: Alignment with organizational values and external ethical frameworks. Evidence that tradeoffs are documented and reviewed by governance stakeholders	☐ 0 ☐ 1 ☐ 2
Data Quality & Integrity Verification	
14. **Validate input data quality:** Check completeness, accuracy, representativeness, and lack of harmful bias. Confirm data preprocessing, annotation consistency, and lineage tracking.	☐ **0** ☐ **1** ☐ **2**
15. Confirm data governance and privacy controls: Validate data minimization, retention limits, secure storage, and controlled access. Ensure personal data protections meet regulatory standards.	☐ **0** ☐ **1** ☐ **2**
16. Test system behavior with real-world and synthetic datasets: Use mixed sources to better evaluate generalizability and detect hidden risks.	☐ **0** ☐ **1** ☐ **2**
Human Oversight, Controls & Safeguards Validation	
17. Validate human-in-the-loop/on-the-loop mechanisms: Confirm operators can intervene, override, or escalate when needed. Ensure oversight tools are intuitive and effective under time-critical conditions.	☐ **0** ☐ **1** ☐ **2**

18. Test operational safeguards and monitoring alerts: Bias alerts, anomaly detection, drift monitoring, and error flagging. Validate that alerts generate timely and actionable signals.	☐ **0** ☐ **1** ☐ **2**
19. Confirm user-facing transparency and explanation features: Ensure any explanations, confidence scores, or flags are understandable and accurate. Test clarity for all user types, including non-technical users.	☐ **0** ☐ **1** ☐ **2**
Risk Assessment, Mitigation & Transparent Reporting	
20. Maintain comprehensive documentation of all TEVV activities: Include technical, operational, ethical, security, and compliance risks. Track risks in a structured registry, Test results, methodologies, benchmarking comparisons, uncertainties, and validation logs.	☐ **0** ☐ **1** ☐ **2**
21. Final readiness verification must be formally approved: Provide deployment recommendations for Leadership, compliance, security, and risk owners must sign off. Approval is based on TEVV results, documentation, and mitigation success.	☐ **0** ☐ **1** ☐ **2**
22. Ensure transparency and traceability: Clear links between requirements → risks → tests → results → approvals. Sufficient detail to reproduce assessments.	☐ **0** ☐ **1** ☐ **2**
23. Implement mitigation strategies and re-test affected components: No high or critical risks may remain unresolved before deployment. Validate that remediations reduce risk to acceptable levels	☐ **0** ☐ **1** ☐ **2**

Total (46) _____

C. Post-Deployment (Monitoring and Audit)

Assessment Item	Score
Performance Monitoring	
1. Continuous performance tracking must be implemented to measure accuracy, latency, reliability, and robustness during real-world operation, using the same—or stricter—metrics defined during pre-deployment TEVV.	☐ 0 ☐ 1 ☐ 2
2. Monitoring systems must detect model performance degradation, data drift, concept drift, or anomalous behavior, triggering alerts or automated responses when performance falls outside approved thresholds.	☐ 0 ☐ 1 ☐ 2
3. Real-time or near-real-time dashboards must provide visibility into system health, model behavior, resource utilization, and operational KPIs to ensure timely detection of issues	☐ 0 ☐ 1 ☐ 2
Risk & Trustworthiness Monitoring	
4. Fairness and bias metrics must be continuously evaluated across demographic groups, contexts, and usage patterns to detect discriminatory outcomes or disproportionate impacts.	☐ 0 ☐ 1 ☐ 2
5. Security monitoring must include detection of adversarial attacks, malicious inputs, model extraction attempts, and abnormal access patterns.	☐ 0 ☐ 1 ☐ 2
6. Safeguards must include automated fallback mechanisms—such as switching to a safe mode, human-in-the-loop intervention, or system shutdown—when harmful or high-risk behavior is detected.	☐ 0 ☐ 1 ☐ 2
7. Monitoring processes must track compliance with trustworthiness characteristics such as Explainability, transparency, resilience, privacy protection, and human oversight effectiveness.	☐ 0 ☐ 1 ☐ 2
Operational & Technical Monitoring	
8. Data quality and integrity must be continuously checked to ensure incoming operational data remains accurate, relevant, and representative, with drift or corruption promptly addressed.	☐ 0 ☐ 1 ☐ 2

9. Model dependencies—including APIs, cloud services, hardware accelerators, and third-party components—must be monitored for availability, stability, and performance impact.	☐ 0 ☐ 1 ☐ 2
10. Logging and observability controls must capture detailed, immutable records of model inputs, outputs, decisions, and system events to support audits, investigations, and root-cause analysis.	☐ 0 ☐ 1 ☐ 2
11. System uptime, reliability, and resilience must be continually assessed, including monitoring of failure rates, error patterns, and system recovery performance.	☐ 0 ☐ 1 ☐ 2
Audit & Compliance Oversight	
12. Periodic technical audits must validate that the system's performance, robustness, and behavior align with approved specifications, regulatory requirements, and internal governance standards.	☐ 0 ☐ 1 ☐ 2
13. Ethical and societal impact audits must assess unintended consequences, user harm, fairness issues, or disproportionate effects emerging from real-world use.	☐ 0 ☐ 1 ☐ 2
14. Privacy audits must verify compliance with data protection laws, retention policies, access controls, and ongoing privacy-by-design obligations.	☐ 0 ☐ 1 ☐ 2
15. Security audits must examine vulnerabilities, threat exposure, access rights, and the effectiveness of security controls and protective mechanisms.	☐ 0 ☐ 1 ☐ 2
16. Audit findings must be formally documented, reportable to governance committees, and transparently communicated to relevant stakeholders.	☐ 0 ☐ 1 ☐ 2
Continuous Improvement & Corrective Action	
17. All monitoring and audit results must feed into a structured improvement cycle that identifies corrective actions, retraining needs, recalibration steps, and system modification requirements.	☐ 0 ☐ 1 ☐ 2
18. Change management procedures must govern updates, ensuring that any retraining, model tuning, or architectural changes undergo TEVV before re-deployment.	☐ 0 ☐ 1 ☐ 2

19. Feedback loops must integrate user input, complaints, incident reports, and insights from affected stakeholders to drive ongoing refinement of the system.	☐ 0 ☐ 1 ☐ 2
20. Emerging risks or patterns identified in post-deployment operation must inform future risk assessments, updates to governance policies, and enhancement of safeguards.	☐ 0 ☐ 1 ☐ 2
21. Lessons learned from incidents, near-misses, or audit findings must be documented and incorporated into organizational knowledge to improve future AI projects.	☐ 0 ☐ 1 ☐ 2

Total (42)_____

TOTAL SCORE of AI Life Cycle Governance (out of 274)	
Design & Development :	**/ 186**
Pre-Deployment Assessment:	**/ 46**
Post-Deployment Assessment:	**/ 42**

D. Stakeholders Engagement

Stakeholder Engagement – Self-Assessment Scoring Tool For Entities Deploying or Supervising AI Systems

Scoring Scale

Score	Meaning	Description
Not Implemented = 0	No evidence	Not performed or undocumented
Partially = 1 Implemented	Some progress	Exists in draft/pilot form or inconsistently applied
Fully Implemented = 2	Operational	Documented, consistent, and reviewed regularly

Maximum Score: 40 points

Readiness Levels:

- **00–15 = Low Readiness**
- **16–30 = Moderate Readiness**
- **31–40 = High Readiness**

SECTION A — Stakeholder Identification, Analysis & Communication (10 points)

Assessment Item	Score (0–2)
1. Stakeholders are formally identified and mapped for each AI use case (citizens, regulators, NGOs, industry, community groups).	☐ 0 ☐ 1 ☐ 2
2. Stakeholder needs, expectations, and potential harms are analyzed and documented.	☐ 0 ☐ 1 ☐ 2
3. AI-enabled public consultation tools (digital participation platforms) are used to improve inclusion and scalability.	☐ 0 ☐ 1 ☐ 2
4. Complaint and appeal channels exist for citizens to report bias or unfair outcomes.	☐ 0 ☐ 1 ☐ 2
5. Reports of harmful patterns are escalated to oversight bodies or authorities.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 10): _____

SECTION B — Collaboration & Shared Accountability (10 points)

Assessment Item	Score
6. Formal collaboration mechanisms exist between government, industry, academia, and civil society.	☐ 0 ☐ 1 ☐ 2
7. Stakeholders are engaged during design, pre-deployment review, and risk assessment.	☐ 0 ☐ 1 ☐ 2
8. Stakeholders can request traceability or explanations for AI-driven decisions.	☐ 0 ☐ 1 ☐ 2
9. Mechanisms allow stakeholders to trigger audits through complaints or collective feedback.	☐ 0 ☐ 1 ☐ 2
10. Collaboration outcomes lead to measurable improvements in system design or policy.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 10): _____

SECTION C — Two-Way Community Communication & Participation (10 points)

Assessment Item	Score
11. Community education tools and awareness programs explain system purpose, risks, and safeguards.	☐ 0 ☐ 1 ☐ 2
12. The organization collects community needs and concerns before deployment.	☐ 0 ☐ 1 ☐ 2
13. Post-deployment feedback mechanisms exist and influence updates.	☐ 0 ☐ 1 ☐ 2
14. Engagement includes diverse internal and external groups to surface harms, benefits, and societal concerns.	☐ 0 ☐ 1 ☐ 2
15. Insights from community engagement are integrated into the risk management process.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 10): _____

SECTION D — Transparency, Disclosure & Responsible Reporting (10 points)

Assessment Item	Score
16. AI system capabilities and limitations are transparently communicated to users.	☐ 0 ☐ 1 ☐ 2
17. Safety and societal risk assessments are shared with relevant stakeholders.	☐ 0 ☐ 1 ☐ 2
18. Acceptable and unacceptable uses are clearly disclosed.	☐ 0 ☐ 1 ☐ 2
19. Channels exist for users to report vulnerabilities, issues, or unexpected behaviors.	☐ 0 ☐ 1 ☐ 2
20. Incentive programs (e.g., bug bounties, contests, prizes) support responsible disclosure.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 10): _____

TOTAL SCORE (out of 40)	
Stakeholder Identification:	**/ 10**
Collaboration:	**/ 10**
Community Communication:	**/ 10**
Transparency & Reporting:	**/ 10**

Overall Stakeholder Engagement Readiness

- **0–15: Low – Limited engagement, high societal risk**
- **16–30: Moderate – Engagement present but uneven**
- **31–40: High – Strong, inclusive, and transparent practices**

3. Society and Sustainability

Society & Sustainability: Self-Assessment Scoring Tool For Government Entities, Developers, & Community Organizations

Scoring Scale

Score	Meaning	Description
Not Implemented = 0	No evidence	Not practiced or undocumented
Partially Implemented = 1	In progress	Exists but inconsistent or ad hoc
Fully Implemented = 2	Operational	Documented, consistent, reviewed regularly

Maximum Score: 60 points

Readiness Levels:

- **0–20 = Low Readiness**
- **21–40 = Moderate Readiness**
- **41–60 = High Readiness**

Section 1: Societal Protection (30 points)

Ensuring AI respects societal values, protects individuals, and promotes responsible use.

A. Purposeful & Responsible Adoption (6 points)

Assessment Item	Score (0–2)
1. AI is adopted only when it adds meaningful value (not trend-driven).	☐ 0 ☐ 1 ☐ 2
2. Human judgment remains the primary basis for decisions.	☐ 0 ☐ 1 ☐ 2
3. AI refines/enhances human reasoning without replacing it.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

B. Protection of Minors (6 points)

Assessment Item	Score
4. AI-enabled services are age-appropriate and designed for children's needs.	☐ 0 ☐ 1 ☐ 2
5. Safeguards exist for children's data (collection, consent, storage).	☐ 0 ☐ 1 ☐ 2
6. Systems align with IEEE 2089 / P7004 child protection standards.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

C. Cultural & Social Alignment (6 points)

Assessment Item	Score
7. AI deployments are reviewed for alignment with Egyptian societal norms.	☐ 0 ☐ 1 ☐ 2
8. Imported AI systems are adapted to local cultural context.	☐ 0 ☐ 1 ☐ 2
9. Cultural and social impact risks are formally assessed.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

D. Transparency & User Rights (6 points)

Assessment Item	Score
10. Users are informed when AI is used in decisions or interactions.	☐ 0 ☐ 1 ☐ 2
11. Clear explanations are provided for AI-driven decisions.	☐ 0 ☐ 1 ☐ 2
12. Users can exercise privacy rights (access/correction/withdrawal).	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

E. Human-Centered Decision Making (6 points)

Assessment Item	Score
13. AI supports rather than replaces decision making in sensitive domains.	☐ 0 ☐ 1 ☐ 2
14. Humans remain accountable for consequential decisions.	☐ 0 ☐ 1 ☐ 2
15. Operators are trained to question, override, and supervise AI outputs.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

F. Responsible Use Boundaries & Privacy-Conscious Behavior (6 points)

Assessment Item	Score
16. AI is used within well-defined and communicated limits (e.g., AI-assist ≠ autonomous).	☐ 0 ☐ 1 ☐ 2
17. Boundary limitations are clearly communicated to users.	☐ 0 ☐ 1 ☐ 2
18. Sensitive data (voice, face, contacts) is shared/used prudently with informed consent.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

Section 2: Environmental Protection (30 points)

Ensuring AI adoption is sustainable, resource-efficient, and environmentally responsible.

A. Right-Sizing & Frugal AI Adoption (6 points)

Assessment Item	Score
19. AI solution complexity matches functional requirements (no excessive models).	☐ 0 ☐ 1 ☐ 2
20. Lightweight / frugal models are evaluated before large models.	☐ 0 ☐ 1 ☐ 2
21. Frugal AI justification is documented when heavier models are chosen.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

B. Computational & Energy Efficiency (6 points)

Assessment Item	Score
22. Energy-efficient training methods are implemented.	☐ 0 ☐ 1 ☐ 2
23.Efficient model architectures (compression, pruning, quantization) are used where feasible.	☐ 0 ☐ 1 ☐ 2
24. Compute-intensive tasks are optimized or scheduled for minimal energy load.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

C. Data Efficiency (6 points)

Assessment Item	Score
25. Data-efficient techniques (transfer learning, synthetic data) are utilized.	☐ 0 ☐ 1 ☐ 2
26. Training datasets are minimized while maintaining performance.	☐ 0 ☐ 1 ☐ 2
27.Data collection avoids unnecessary or excessive data accumulation.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

D. Environmental Impact Assessment (6 points)

Assessment Item	Score
28. All AI proposals include estimated compute needs and energy use.	2 ☐ 1 ☐ 0 ☐
29. Carbon footprint or environmental impact is assessed for high-impact systems.	2 ☐ 1 ☐ 0 ☐
30. Environmental risks are reviewed before approving system deployment.	2 ☐ 1 ☐ 0 ☐

Subtotal (max 6): _____

E. Sustainable Infrastructure & Monitoring (6 points)

Assessment Item	Score
31. AI is deployed on energy-efficient or renewable-powered infrastructure.	☐ 0 ☐ 1 ☐ 2
32. Resource usage (compute, storage, electricity) is monitored continuously.	☐ 0 ☐ 1 ☐ 2
33. Sustainability performance (e.g., cost, energy efficiency) is reviewed periodically.	☐ 0 ☐ 1 ☐ 2

Subtotal (max 6): _____

TOTAL SCORE (out of 60)	
Societal Protection:	**/ 30**
Environmental Protection:	**/ 30**

Overall Society & Sustainability Readiness Level

- **0–20:** Low Readiness — High societal & environmental risk
- **21–40:** Moderate Readiness — Improvements needed
- **41–60:** High Readiness — Strong safeguards in place

Responsible AI system readiness level

The total combined score for the 4 pillars is 370 points, distributed as follows:

#	Pillar	Max
2	AI Life Cycle Governance	270
3	Stakeholders Engagement	40
4	Societal & Sustainability	60
	Overall Score	370

Responsible AI system Readiness Levels (based on total score of all pillars)

- **0–148:** Low Readiness — High risk; major shortcomings in governance and system trustworthiness.
- **149–259:** Moderate Readiness — Partial alignment; targeted improvements required prior to scaled deployment.
- **260–370:** High Readiness — Strong responsible AI practices embedded; system is reliable, trustworthy, and well-governed.

Annex 2: Trustworthy AI Tools: Ensuring Fair, Explainable, and Safe AI Systems

1. Introduction

As artificial intelligence increasingly powers critical applications, establishing trustworthiness in AI systems is essential for widespread adoption. A trustworthy AI model is one that ensures fairness by minimizing bias and discrimination; explainability by making its decisions understandable; robustness by resisting adversarial attacks and failures; safety by preventing harmful outputs; transparency and accountability through traceability and auditability; and respect for user privacy. To meet these criteria, researchers and organizations have developed a diverse set of tools and frameworks. This report reviews the technical resources available for assessing and enhancing AI trustworthiness, covering both traditional machine learning techniques and modern generative AI like large language models (LLMs). It includes open-source libraries as well as commercial solutions, with an emphasis on deploying these tools in enterprise environments.

2. Trustworthiness in Traditional Machine Learning

Traditional ML models (e.g. regression, decision trees, neural networks on structured data) require tools to evaluate and improve various trust aspects before and during deployment. Key categories include:

Fairness and Bias Mitigation

A variety of AI fairness toolkits assist in identifying and reducing unwanted biases in data and models:

- **IBM AI Fairness 360 (AIF360):**

  Open-source library providing numerous fairness metrics for datasets/models and bias mitigation algorithms.

- **Microsoft Fairlearn:**

  Toolkit offering fairness metrics across demographic subgroups and algorithms to improve parity in model outcomes.

- **Aequitas:**

  An open-source audit toolkit focused on detecting discrimination or unequal error rates in machine learning models.

- **Themis-ML:**

  An earlier library, extending scikit-learn, that introduced fairness metrics **and** enabled training under fairness constraints.

- **PandasAI:**

  A newer tool leveraging LLMs to help analysts find biases in tabular data through natural language queries.

- **TensorFlow Extended:**

  incorporate TensorFlow Model Analysis (TFMA) with Fairness Indicators to calculate metrics for different data segments and visualize bias issues on dashboards before deployment

- **Amazon SageMaker**

  Clarify and IBM Watson OpenScale offer continuous bias monitoring in production.

Explainability and Interpretability

To trust an AI's decisions, users and regulators often need explanations. A broad set of explainable AI (XAI) tools provide insight into model behavior:

- InterpretML trains inherently interpretable models (Explainable Boosting Machines) and explains black-box models.
- LIME and SHAP provide local and global explanations via perturbation and Shapley values.
- ELI5 explains scikit-learn models by showing weights and feature importances.
- Captum explains neural network decisions with saliency maps.
- Grad-CAM generates heatmaps highlighting image areas influencing predictions.

- Google's Facets and What-If Tool (WIT) enable interactive exploration of datasets and model predictions.

Enterprise solutions such as IBM Watson OpenScale, Fiddler AI, and H2O Driverless AI generate explanation dashboards, reason codes, and bias summaries for regulators and business users.

Robustness and Security (Adversarial ML)

AI models can be vulnerable to adversarial attacks and input perturbations that cause erratic behavior, undermining reliability. Tools for robustness testing help evaluate and improve model resilience:

- IBM Adversarial Robustness Toolbox (ART) and CleverHans create adversarial examples (e.g., FGSM, DeepFool) and provide defenses like adversarial training.
- Alibi Detect and Microsoft Counterfit simulate attacks and monitor sensitivity to data shifts.
- Neural Cleanse identifies hidden backdoors by reverse-engineering trigger patterns.
- The MITRE ATLASdatabase catalogs known adversarial techniques and countermeasures.

Uncertainty and Reliability

Reliable AI requires awareness of prediction uncertainty:

- IBM Uncertainty Quantification 360 (UQ360) offers methods such as conformal prediction and Monte Carlo dropout to estimate confidence.
- **Microsoft** Responsible AI Toolbox detects conditions where models perform poorly (e.g., specific subpopulations).

Integrating uncertainty measures prevents overconfidence in low-certainty predictions and supports human-in-the-loop decision-making.

Privacy-Preserving ML

AI systems handling personal data must meet strict privacy requirements:

- **Differential Privacy:** Google's Differential Privacy library, TensorFlow Privacy and PyTorch's Opacus add statistical noise to protect individuals.
- **Federated Learning:** Platforms like FATE and Flower train models collaboratively without data sharing.
- **Homomorphic Encryption:** Microsoft SEAL and Concrete-ML enable encrypted inference on sensitive data.
- **Secure Computation:** PySyft applies cryptographic protocols for decentralized model training.
- ML Privacy Meter audits trained models for membership inference vulnerabilities.

These tools support GDPR and HIPAA compliance by ensuring models and data pipelines preserve confidentiality throughout the AI lifecycle.

Transparency, Accountability, and Governance

Beyond algorithms, governance frameworks ensure responsible deployment:

- Google's Model Card Toolkit and IBM's AI FactSheets document model design, datasets, metrics, and limitations.
- Credo AI and Holistic AI manage model inventories, enforce governance policies, and automate regulatory compliance (EU AI Act, ISO 42001).
- IBM Watson OpenScale, Arize AI, WhyLabs, Monte Carlo, Coralgix, Arthur, Fiddler, and TruEra continuously monitor production models for drift, bias, and audit readiness.

These systems combine documentation, monitoring, and compliance workflows to maintain accountability throughout the model lifecycle.

3. Trustworthiness in LLMs and Generative AI

Large Language Models (LLMs) and generative AI systems, such as GPT, Claude, and image generators, have attracted significant enterprise interest due to their powerful capabilities. However, they introduce new trust challenges including hallucinated outputs (confident but incorrect information), toxic or biased language, vulnerability to prompt injection attacks, and risks of sensitive data leakage within prompts or responses. Because generative AI engages in open-ended and dynamic interactions, traditional static evaluation methods are insufficient. This has driven the creation of specialized tools and frameworks focused on LLM trustworthiness, often referred to as "Responsible GenAI" or "AI Governance for LLMs."

Evaluation Frameworks for Generative Models

Evaluating large language models requires testing for factuality, safety, and bias:

- Stanford's Holistic Evaluation of Language Models benchmarks LLMs across diverse metrics, accuracy, calibration, robustness, and toxicity.
- OpenAI Evals framework enables custom automated tests for content validity and refusal behavior.
- NVIDIA's Garak,Microsoft's PyRIT, and Prompt Fuzzer conduct automated red-teaming to expose vulnerabilities such as prompt injection or data leakage.
- **Content moderation APIs** (OpenAI's content filter and Google's Perspective API) detect harmful or policy-violating outputs.
- Project Moonshot, developed by Singapore's AI Verify Foundation, integrates benchmarking with adversarial safety exams, functioning as a "crash test" for full AI applications.

These frameworks allow systematic and reproducible safety evaluations prior to deployment.

Alignment and Control Tools

Alignment ensures LLMs follow desired behavior:

- Guardrails AI enforces validation rules and output schemas to block PII, profanity, or factual errors.
- OpenAI Function Calling and structured output enforcement constrain models to safe formats.
- Microsoft's Prompt Flow and the LangChain framework manage multi-step prompt workflows with moderation and validation.
- TruLens and Constitutional AI establish feedback loops for model self-correction and alignment tuning.

Such runtime "safety layers" complement training-time alignment and maintain behavioral control.

Monitoring and Governance for LLMs

Responsible LLM deployment requires continuous oversight:

- Lakera Guard acts as a GenAI firewall, scanning prompts and responses for injections, data leaks, or toxic content.
- Arthur and Coralogix provide real-time monitoring for hallucinations and safety violations.
- CalypsoAI enforces access control and policy compliance, preventing unauthorized model use or data exposure.
- Credo AI, Monitaur, and Mozart Data extend governance to generative systems, tracking usage, data provenance, and regulatory adherence.

These solutions deliver transparency, risk management, and policy enforcement across enterprise LLM deployments.

4. Enterprise Deployment Considerations

For effective implementation, trust tools must integrate across the AI lifecycle:

- **Lifecycle Integration:** Embed fairness, robustness, and explainability checks into development, CI/CD pipelines, and post-deployment monitoring using frameworks such as TensorFlow Model Analysis, OpenAI Evals framework, or IBM Adversarial Robustness Toolbox.
- **Performance vs. Thoroughness:** Combine lightweight real-time filters (e.g., content moderation) with periodic deep audits (e.g., bias or adversarial testing).
- **Cross-Functional Governance:** Use platforms with business-friendly dashboards (Credo AI, IBM Watson OpenScale) to involve compliance and domain experts.
- **Regulatory Alignment:** Leverage tools that support the EU AI Act and NIST AI RMF through automated documentation and continuous risk monitoring.
- **Security and Privacy:** Employ encrypted logging and on-premise deployment where necessary (CalypsoAI, Azure Confidential ML) to protect sensitive data.

This integrated, multidisciplinary approach ensures technical and regulatory robustness.

The ecosystem of trustworthy AI tools now covers every stage of the ML and LLM lifecycle. Traditional ML toolkits, such as, AIF360, SHAP, ART, TFMA, and others, address fairness, explainability, robustness, and privacy. New frameworks, such as, HELM, Guardrails AI, Moonshot, and Lakera Guard, extend these principles to generative AI, ensuring safe and aligned interactions. Enterprises adopting these technologies can achieve not only accurate and efficient AI but also systems that are fair, transparent, robust, and compliant by design, a foundation for sustainable, responsible AI strategy.

Annex 3: International Standards

P1: Governance, P2: Infrastructure, P3: Technology, P4: Data, P5: Talents, P6: Ecosystem

A. ISO Standards: Living List of Technical and Foundational Standards, Technical Specifications, and Technical Reports of relevance to the Egyptian AI Governance Framework and the second edition of the Egyptian National AI Strategy						
Name of standard/ TS/ TR	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
ISO/IEC 27001: 2022 Information security management systems- requirements	X	X				
ISO/IEC TS 8200:2024 AI- Controllability of automated artificial intelligence systems	X					
ISO/IEC 8183:2023 Information Technology —Artificial intelligence — Data life cycle framework	X			X		
ISO/IEC TR 24029-1:2021 Artificial Intelligence (AI) — Assessment of the robustness of neural networks — Part 1: Overview			X			
ISO/IEC 24029-2:2023 Artificial intelligence (AI) — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods			X			
ISO/IEC TR 24030:2024 Information technology — Artificial intelligence (AI) — Use cases			X			
ISO/IEC TR 24368:2022 Information technology — Artificial intelligence — Overview of ethical and societal concerns	X					
ISO/IEC TR 24372:2021 Information technology — Artificial intelligence (AI) — Overview of computational approaches for AI systems			X			
ISO/ IEC 24668:2022 Information technology — Artificial intelligence — Process management framework for big data analytics		X		X	X	

ISO/IEC TR 24027:2021 Bias in AI systems and AI aided decision making	X		X	X		
ISO/IEC 5259 series Data quality for analytics and machine learning ML				X		
ISO/IEC 38505 series Governance of data	X			X		
ISO/IEC 29100:2024 Security techniques	X			X		
ISO/IEC 27018:2025 Information Security and privacy protection- Guidelines for protection of personally identifiable Information (IPP)in public clouds acting as PII processors	X	X		X		
ISO/IEC 23894: 2023 AI Guidance on risk management	X	X		X		
ISO/IEC 38507:2022 Governance implications of the use of artificial intelligence by organizations	X					X
ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system	X		X			X
ISO/IEC 42005:2025 Information Technology- Artificial Intelligence- AI system impact assessment	X					X
ISO/IEC TR 24368:2022 AI overview of ethical and societal concerns	X					
ISO/IEC TR 24028:2020 Overview of trustworthiness in artificial intelligence	X					
ISO/IEC 22989:2022 AI Concepts and terminology	X					

ISO/IEC 23053:2022 Framework for AI systems using Machine learning			X			
ISO/IEC 5338:2023 AI Systems life cycle processes	X					
ISO/IEC 19941:2017 Cloud computing- interoperability and portability		X		X		X
ISO/ IEC TS 4213: 2022 Information technology- Artificial intelligence - Assessment of machine learning classification performance			X			
ISO/IEC 5259-1: 2024 Artificial Intelligence- Data quality for analytics and machine learning (ML)- Part 1: Overview, terminology and examples		X		X		X
ISO/IEC TS 12791: 2024 Information technology — Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks	X					
ISO/IEC TR 17903:2024 Information technology — Artificial intelligence — Overview of machine learning computing devices			X			
ISO/IEC 17021-1:2015 Conformity Assessment- Requirements for bodies providing audit and Certification of Management systems	X					
ISO/IEC 42006:2025 Information technology — Artificial intelligence — Requirements for bodies providing audit and certification of artificial intelligence management systems	X					
ISO/ IEC 5339: 2024 Information technology — Artificial intelligence — Guidance for AI applications	X		X			
ISO/IEC 5392:2024 Information technology — Artificial intelligence — Reference architecture of knowledge engineering			X			

ISO/IEC TR 5469:2024 Artificial intelligence — Functional safety and AI systems	X		X			
ISO/IEC TR 20226:2025 Information technology — Artificial intelligence — Environmental sustainability aspects of AI systems	X					
ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary				X		
ISO/IEC 20547-1:2020 Information technology — Big data reference architecture — Part 1: Framework and application process				X		
ISO/IEC TR 20547-2:2018 Information technology — Big data reference architecture — Part 2: Use cases and derived requirements				X		
ISO/IEC TR 20547-5:2018 Information technology — Big data reference architecture — Part 5: Standards roadmap				X		
ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology	X					
ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems (Using Machine Learning (ML			X			
ISO/IEC 23894:2023 Information technology — Artifi cial intelligence — Guidance on risk management	X					
ISO/IEC 25059:2023 Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems	X					

References:

1. ISO standards Portal: https://www.iso.org/standards.html
2. Ai for Good data base: https://aiforgood.itu.int/ai-standards-exchange/ai-standards-exchange-database/
3. EOS National Committee of SC42; Report by engineer Mohamed Khairy.
4. Links to the above mentioned standards can be found at the Egyptian NAIC portal: **www.aic.gov.eg**

B. ITU Standards: Living List of ITU standards/ pre standards/ Technical Reportsand Reports of relevance to the Egyptian AI Governance Framework						
Standard/ Report/ Technical Report	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**
ITU-T Y.3170-series – Machine learning in future (networks including IMT-2020: Use cases (2019		X	X			
Architectural framework for machine learning in future networks including IMT-2020 ITU-T Y.3172 (06/2019) ML for IMT2020 (FG-ML5G) FG ML5G was active from January 2018 until July 2020		X	X			
ITU-T Q.5007 (12/2023) Signaling architecture for microservices based intelligent edge computing		X	X			
Shaping ethics, regulation and standardization in AI for health ITU-WHO Focus Group on AI for Health Final (Report (2025			X			
Regulatory considerations on artificial intelligence for health (FG-AI4H DEL02 (09/2022			X			
Digital Agriculture: A Standards Snapshot ITU/FAO Focus Group on AI and IoT for Digital Agriculture ((2025			X			
ITU-T Focus Group Technical Report (FG-AI4A WG-ELR): Ethical, legal and regulatory considerations relating to the use of AI (for agriculture: A European perspective (03/2024			X			
(ITU-T F.742.1 (12/2022 Requirements for smart class systems based on artificial intelligence SERIES F: Non-telephone telecommunication ser-vices Multimedia services		X	X			

(ITU-T F.743.27 (01/2025 Requirements and framework of intelligent video surveillance platform for power grid infrastructure SERIES F: Non-telephone telecommunication services Multimedia services		X	X			
(ITU-T F.746.15 (12/2022 Requirements for smart broadband network gateway in multimedia content transmission SERIES F: Non-telephone telecommunication services Multimedia services		X	X			
(ITU-T F.748.21 (12/2022 Requirements and framework for feature-based distributed intelligent systems SERIES F: Non-telephone telecommunication services Multimedia services		X	X			
(ITU-T F.748.22 (09/2023 Functional architecture for feature-based distributed intelligent systems		X	X			
(ITU-T Y.3533 (12/2023 Cloud computing – Functional requirements for robotics as a service SERIES Y: Global information infrastructure, Internet protocol aspects, next-generation networks, Internet of Things and smart cities		X	X			
AI Standards for Global Impact: From Governance to action	X	X	X	X	X	X
Digital Agriculture: A Standards Snapshot ITU/FAO Focus Group on AI and IoT for Digital Agriculture			X			

References:

Ai 4 Good data base:
https://aiforgood.itu.int/ai-standards-exchange/ai-standards-exchange-database/

Hyperlinks to ITU Standards, Reports, and Technical Reports mentioned above are found on the Egyptian NAIC portal: www.aic.gov.eg

C. Living List of IEEE Standards relevant to the Egyptian AI Governance Framework and related to the 2nd edition of the Egyptian AI Strategy						
IEEE Standard	**Governance**	**Infrastructure**	**Technology**	**Data**	**Talents**	**Ecosystem**
IEEE 7000-2021 – Model Process for Addressing Ethical Concerns During System Design	✓					
IEEE 7001-2022 – Transparency of Autonomous Systems			✓			
IEEE 7002-2022 – Data Privacy Process				✓		
IEEE 7003-2024 – Algorithmic Bias Considerations			✓			
IEEE P7004 – Child and Student Data Governance				✓		
IEEE 7005-2021 – Transparent Employer Data Governance				✓		
IEEE 7007-2021 – Ontological Standard for Ethically Driven Robotics and Automation Systems			✓			
IEEE 7008 (P7008) – Ethically Driven Nudging for Robotic, Intelligent and Autonomous Systems			✓			
IEEE 7009-2024 – Fail-Safe Design of Autonomous and Semi-Autonomous Systems		✓				
IEEE P7009.1 – Safety Management for Autonomous Sys-(tems (Anomalous Behavior Interventions		✓				

IEEE 7010-2020 – Assessing Impact of AI on Human Well-Being						✓
IEEE 2089-2021 – Age-Appropriate Digital Services Framework (5Rights Principles				✓		
IEEE 2801-2022 – Quality Management of Datasets for Medical AI				✓		
IEEE 3119-2025 – Procurement of AI and Automated Decision Systems	✓					
IEEE 3410-2025 – Guide for Large-Scale AI Models in Financial Risk Management			✓			
IEEE 2840-2025 – Responsible AI Licensing	✓					
IEEE P2863 – Organizational Governance of Artificial Intelligence	✓					
IEEE P3396 – Defining and Evaluating AI Risk, Safety, Trustworthiness, and Responsibility (Framework	✓					
IEEE P3511 – Risk Management for Generative AI Systems	✓					
IEEE P7999 – Integrating Organizational AI Ethics Oversight ((Framework	✓					

References:

IEEE Portal: https://IEEE.org

Ai for good data base:
https://aiforgood.itu.int/ai-standards-exchange/ai-standards-exchange-database/

Hyperlinks to IEEE Standards mentioned above are found on the Egyptian NAIC portal:
www.aic.gov.eg

Annex 4: Acronyms

Roles & Stakeholders

- **G:** Government / Governmental Agencies
- **E:** Enterprise (Private-sector developers and service providers)
- **C**: Community (Individuals, groups, or civic entities)
- **PD**: Provider/Developer
- **BU**: Beneficiary/User
- **CAIO**: Chief AI Officer

Technical & Operational Terms

- **AI**: Artificial Intelligence
- **GenAI**: Generative AI
- **ML**: Machine Learning
- **LLM**: Large Language Model
- **TEVV**: Test, Evaluate, Verify, and Validate
- **XAI**: Explainable AI
- **HITL**: Human-in-the-loop
- **HOTL**: Human-over-the-loop (or Human-on-the-loop)
- **HOOTL**: Human-out-of-the-loop
- **KPIs**: Key Performance Indicators
- **API**: Application Programming Interface
- **PII**: Personally Identifiable Information
- **SQuaRE**: Systems and software Quality Requirements and Evaluation (ISO standard series)

Organizations & Authorities

- **MCIT**: Ministry of Communications and Information Technology
- **NCAI**: National Council for Artificial Intelligence, Quantum Computing, and Emerging Technologies
- **EGCERT**: Egyptian Computer Emergency Response Team
- **SECC-ITIDA**: Software Engineering Competence Center - Information Technology Industry Development Agency
- **ECRAI**: Egyptian Council for Responsible AI (Contextually implied regulatory/ audit entity)

International Standards & Bodies

- **ISO**: International Organization for Standardization
- **IEC**: International Electrotechnical Commission
- **ITU**: International Telecommunication Union
- **IEEE**: Institute of Electrical and Electronics Engineers
- **OECD**: Organization for Economic Co-operation and Development
- **UNESCO**: United Nations Educational, Scientific and Cultural Organization
- **GPAI**: Global Partnership on Artificial Intelligence
- **NIST**: National Institute of Standards and Technology
- **WHO**: World Health Organization
- **FAO**: Food and Agriculture Organization

Laws, Regulations & Frameworks

- **PDPL**: Personal Data Protection Law (Law No. 151 of 2020)
- **GDPR**: General Data Protection Regulation (European Union)
- **HIPAA**: Health Insurance Portability and Accountability Act (USA)
- **HUDERIA**: Human Rights Education for Legal Professionals in AI (Council of Europe framework)