DRAFT KENYA STANDARD

DKS 3007:2024

**ICS 01.040.35;
35.020**

First Edition

Information technology — Artificial Intelligence — Code of Practice for AI Applications

Kenya Bureau of Standards
Standards for Quality life

TECHNICAL COMMITTEE REPRESENTATION

The following organizations were represented on the Technical Committee:

Kenya Airports Authority (KAA)
Tech Innovators Network
IDEAZ Software
Kenya Engineering Technology Registration Board (KETRB)
Daystar University
Impulse Innovations
ISACA Kenya Chapter
Jipee Ajira Limited
Kenya National Library Services
MultiMedial University
Muhoroni Sugar Company Limited
National Industrial Training Institute
Office of the President (National Economic and Social Council)
Social Enterprise Society of Kenya (SESOK)
Kenya Bureau of Standards — SecretariatKenya Bureau of Standards — Secretariat

REVISION OF KENYA STANDARDS

In order to keep abreast of progress in industry, Kenya Standards shall be regularly reviewed. Suggestions for improvements to published standards, addressed to the Managing Director, Kenya Bureau of Standards, are welcome.

DRAFT KENYA STANDARD

DKS 3007:2024

ICS **01.040.35;**
35.020

First Edition

Information technology — Artificial Intelligence — Code of practice for AI Applications

Kenya Bureau of Standards, Popo Road, Off Mombasa Road,
P.O. Box 54974 - 00200, Nairobi, Kenya

☎ +254 020 6948000, + 254 722202137, + 254 734600471

✉ info@kebs.org

@KEBS_ke

kenya bureau of standards (kebs)

Foreword

This Kenya Standard was prepared by the Software Engineering, IT Service Management, IT Governance and Artificial Intelligence Technical Committee under the guidance of the Standards Projects Committee, and it is in accordance with the procedures of the Kenya Bureau of Standards.

During the preparation of this standard, reference was made to the following document (s):

KS ISO/IEC 5339, Information technology — Artificial intelligence — Guidance for AI applications

KS ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system

KS ISO/IEC 5338:2023, Information technology — Artificial intelligence — AI system life cycle processes

NIST AI 100-1, Artificial Intelligence Risk Management Framework (AI RMF 1.0)

EU's Artificial Intelligence Act, March 2024

Acknowledgement is hereby made for the assistance derived from these sources.

Introduction

Artificial intelligence (AI) is increasingly applied across all sectors utilizing information technology and is expected to be one of the main economic drivers. A consequence of this trend is that certain applications can give rise to societal challenges over the coming years.

Artificial intelligence (AI) systems have the potential to create incremental changes and achieve new levels of performance and capability in domains such as agriculture, transportation, fintech, education, energy, healthcare and manufacturing. However, the potential risks related to lack of trustworthiness can impact AI implementations and their acceptance. AI applications can involve and impact many stakeholders, including individuals, organizations and society as a whole. The impact of AI applications can evolve over time, in some cases due to the nature of the underlying data or legal environment. The stakeholders should be made aware of their roles and responsibilities in their engagement.

AI can introduce substantial risks and uncertainties. Professionals, researchers, regulators and individuals need to be aware of the ethical and societal concerns associated with AI systems and applications. Potential ethical concerns in AI are wide ranging.

Examples of ethical and societal concerns in AI include privacy and security breaches to discriminatory outcomes and impact on human autonomy. Sources of ethical and societal concerns include but are not limited to:

— unauthorized means or measures of collection, processing or disclosing personal data;

— the procurement and use of biased, inaccurate or otherwise non-representative training data;

— opaque machine learning (ML) decision-making or insufficient documentation, commonly referred to as lack of explainability;

— lack of traceability;

— insufficient understanding of the social impacts of technology post-deployment.

AI can operate unfairly particularly when trained on biased or inappropriate data or where the model or algorithm is not fit-for-purpose.

The values embedded in algorithms, as well as the choice of problems AI systems and applications are used for to address, can be intentionally or inadvertently shaped by developers' and stakeholders' own worldviews and cognitive bias.

This document contains guidance for AI applications based on a common framework, to provide multiple macro-level perspectives. It also incorporates AI characteristics and non-functional characteristics such as trustworthiness and risk management. The guidance can be used by standards developers, application developers and other interested parties. Since AI applications can differ from non-AI software applications due to their continuously evolving nature and aspects of trustworthiness, all stakeholders should be made aware of AI-specific characteristics.

.

Information technology — Artificial Intelligence — Code of Practice for AI applications

1 Scope

This document provides a set of recommendations intended to help the organization develop, provide, or use AI systems responsibly in pursuing its objectives and meet applicable requirements, obligations related to interested parties and expectations from them. It includes the following:

— approaches to establish trust in AI systems through transparency, explainability, controllability, etc.

— engineering pitfalls and typical associated threats and risks to AI systems, along with possible mitigation techniques and methods; and

— approaches to assess and achieve availability, resiliency, reliability, accuracy, safety, security and privacy of AI systems

This document is applicable to any organization, regardless of size, type and nature, that provides or uses products or services that utilize AI systems.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

KS ISO/IEC 22989:2022, Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

KS ISO/IEC 25059, Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems

KS ISO/IEC TR 24368, Information technology — Artificial intelligence — Overview of ethical and societal concerns

KS ISO/IEC 23894, Information technology — Artificial intelligence — Guidance on risk management

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 22989 and the following apply..

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1 bias

systematic difference in treatment of certain objects, people, or groups in comparison to other

3.7

fairness treatment,

behaviour or outcomes that respect established facts, societal norms and beliefs and are not determined or affected by favouritism or unjust discrimination

4 Characteristics and Processes of Artificial Intelligence Systems

4.1 An AI application can be distinguished from a non-AI application by its possession of one or more of the following functional characteristics. The AI stakeholders described here play one or more different roles and sub-roles in various stages of the AI system life cycle. The name of the stakeholder is also indicative of its role or sub-role as described in KS ISO/IEC 22989:2022, 5.19

	AI Characteristics	Processes	Stakeholders/Actors	Roles and responsibilities (Annex A)
4.1.1	Built with the capabilities of an AI system that implements a model to acquire information and processes with or without human intervention by algorithm or programming.	**AI model and development, AI system** The AI model can be developed from different technologies, such as neural networks, decision trees, Bayesian networks, logic sentences and ontologies. These models are used to make predictions or to compute decisions to support the functions of the AI system.	Data Providers AI Developers AI Producers AI Customer	**A data provider** (Who) is an organization or entity that is concerned with providing data used by AI products or services. A data provider either collects or prepares data (What), or both for use by the AI producer's AI model. The data provider can be a partner of the AI producer. The role of a data provider is usually centred around pre-deployment stages (When). In certain circumstances, such as where the AI system employs machine learning models, the data provider can also be involved in the post-deployment stages to collect and prepare data for continuous validation (When). **An AI developer** (Who) is an organization or entity that is concerned with the development of AI products and services for the producer. The roles can include model and system design, development, implementation, verification and validation (What) in the pre-deployment stages of the AI system life cycle (When). An individual AI developer can be a member of the producer's organization or a contractor or partner **An AI producer** (Who) is an organization or entity that designs, develops, tests and deploys products or services that use one or more AI systems. The AI producer takes on these roles as part of its organization's objective (Why, e.g. profit as well as value creation for its customers). These roles span the whole AI system life cycle (When) and include management decisions about the inception and termination or retirement of the AI system. **An AI customer** (Who) is an organization or entity that uses an AI product or service either directly or by its provision to AI users. There is a business relationship between an AI application provider (see 5.3.2.6) and an AI customer, e.g. engagement, product purchase or service subscription. The customers' role spans the AI system life cycle (When) since they create the demand, realize the value and sustain the viability of the AI product (Why). They are often consulted by the AI producer during the inception to determine requirements and participate in the verification and validation, deployment, operation and monitoring, retirement stages of the AI system

	AI Characteristics	Processes	Stakeholders/Actors	Roles and responsibilities (Annex A)
				life cycle. **AI partner** - An AI partner is an organization or entity that provides services to the AI producer and AI application provider as part of a business relationship.
4.1.2	Applies optimizations or inferences made with the model to augment decisions, predictions or recommendations in a timely manner to meet specific objectives.	**AI application, AI-augmented decision-making** The AI system capabilities are applied to a decision-making environment in a particular domain, including agriculture, transportation, fintech, education, energy, healthcare, manufacturing and many others.	**Internal and external Application providers** **Regulators and policy makers**	**AI application provider** is an organization or entity that provides products or services that uses one or more AI systems. In the AI application context, an AI application provider (Who) is an organization or entity that provides the capabilities from an AI system (such as reasoning and decision-making) in the form of an AI application (What) as a product or service (How) to internal or external customers as described in KS ISO/IEC 22989:2022 **A regulator** (Who) is an authority in the locality where the AI application is deployed and operated, and which has jurisdiction governing the use of AI technology based on existing legal requirements. Even though compliance to legal requirements is assessed by regulators in the deployment, operation and monitoring stages, the AI provider and other early-stage stakeholders should identify applicable risks and regulation and provide solutions to avoid barriers to achieve original objectives. **A policy maker** (Who) is an authority in the locality where the AI application is deployed and operated that sets the legal requirements governing the use of AI technology.
4.1.3	Updates and improvements made to the model, system or application by evaluation of interaction outcomes.	**Continuous validation**	**Internal and external AI customers/Users** **Community**	**An AI user** (Who) is an organization or entity that uses AI products or services. An AI user can be an individual from the community (Who) or a member of the customer organization or entity. A customer can also be a user. An AI user does not have to be an AI customer [i.e. has a business relationship with the AI application provider (see 5.3.2.6)]. An AI user's role is usually centred around the operation and monitoring stage of the AI system life cycle (When) to realize value from use of the AI product or service (Why) **Community** - The use of AI technology can have impacts beyond the individual customer and user and affect other community members (Who) (e.g. consumers, family, neighbours, work colleagues, social circle, affiliates).

5 AI application non-functional characteristics and consideration

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
5.1 Trustworthiness - Trustworthiness is a non-functional and essential characteristic of an AI system. It refers to the characteristic that signifies that the system meets the expectation of its stakeholders in a verifiable way; as well as expressing its quality as being dependable and reliable.	**5.1.1 AI robustness —** AI robustness is the ability of an AI system to maintain its level of performance, as intended by its developers, and required by its customers and users, under any circumstances	**i)** Use a wide variety of testing methods across a spectrum of tasks and contexts prior to deployment to measure performance and ensure robustness. ii) Employ adversarial testing (i.e. red-teaming) to identify vulnerabilities. iii) Perform an assessment of cyber-security risk and implement proportionate measures to mitigate risks, including with regard to data poisoning. iv) Perform benchmarking to measure the model's performance against recognized standards.	▪ **System Design and Architecture Documentation** ○ **Design Specifications** - AI system's architecture, including components, interactions, and dependencies. ○ **Model Architecture**: machine learning model, its layers, and parameters. ▪ **Data Collection and Preprocessing Records** ○ **Data Sources**: Document information about data sources, quality, and any preprocessing steps applied. ○ **Data Augmentation**: Record techniques used for data augmentation to enhance robustness. ▪ **Model Training and Hyperparameters:** ○ Training Settings: Document training configurations, optimization algorithms, and learning rates. ○ **Record hyperparameter values used during model training.** ▪ **Testing and Evaluation Records** ○ **Adversarial Testing**: Document results from adversarial testing (e.g., FGSM, PGD) to assess robustness. ○ **Adversarial Testing**: Document results from adversarial testing (e.g., FGSM, PGD) to assess robustness. ▪ **Error Analysis and Failure Modes** ○ **Failure Cases:** Document instances where the model failed or exhibited vulnerabilities.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
			○ **Failure Cases:** Document instances where the model failed or exhibited vulnerabilities. ▪ **Model Updates and Maintenance** ○ **Version Control:** Maintain records of model versions and updates. ○ **Retraining Cycles:** Document retraining schedules and improvements made.
	5.1.2 AI reliability — AI reliability is the ability of an AI system or any of its subcomponents to perform its required functions under stated conditions for a specific period of time.	i) **Data Quality and Preprocessing:** collecting high-quality, diverse, and representative data. Cleanse the data to remove noise, inconsistencies, and outliers. Augment the dataset if necessary to enhance its and robustness. ii) **Cross-Validation**: Cross-validation helps developers to detect overfitting (the model memorizing the training data) and assess the model's ability to generalize to new data. iii) **Hyperparameter Tuning:** Systematically tuning hyperparameters and evaluating the model's performance, to enhance the accuracy and robustness of AI models. iv) **Model Evaluation Metrics:** Use Model evaluation metrics to assess the performance of AI models quantitatively evaluate the performance of AI models and make informed decisions regarding their deployment	▪ **Data Collection and Preprocessing** ○ **Data Sources**: Record information about data sources, including their quality, diversity, and representativeness. ○ **Data Preprocessing Steps**: Document data cleaning, augmentation, and any transformations applied to the data. ▪ **Model Development and Training:** ○ **Model Architecture**: Detailed description of the chosen model architecture and hyperparameters. ○ **Training Process**: Record training settings, convergence criteria, and any fine-tuning steps. ○ **Validation and Testing**: Document validation metrics, test results, and any model adjustments. ▪ **Model Explainability and Interpretability**: ○ **Explainability Techniques**: Describe how the model's decisions are explained (e.g., SHAP values, LIME). ○ **Interpretability Insights**: Record insights gained from interpreting model behavior. ▪ **Testing and Validation**: ○ **Test Plans**: Detailed plans for testing the AI system, including test cases and expected outcomes. ○ **Validation Reports**: Document results from holdout validation, cross-validation, and A/B testing.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
			▪ **Monitoring and Maintenance**: ○ **Monitoring Protocols**: Specify how the system will be monitored in production. ○ **Maintenance Logs**: Record updates, retraining cycles, and any adjustments made over time. ▪ **Risk Assessment and Mitigation**: ○ **Risk Register**: Identify potential risks (e.g., biases, adversarial attacks) and mitigation strategies. ○ **Ethical Considerations**: Document ethical guidelines followed during development
	5.1.3 AI resilience — AI resilience is the ability of an AI system to recover operational condition quickly following a fault or disruptive incident. Some fault tolerant systems can operate continuously after such an incident, albeit with degraded capabilities.	i) **Governance** - A strong governance structure with Clear Policies and Acceptable Use. Define acceptable boundaries and constraints to prevent misuse or unintended consequences. ii) **Observability-** identifying and cataloguing every AI system or technology deployed within the organization. It's critical to have a clear view of your entire AI ecosystem to monitor activities and detect potential threats in real time. **iii) Regular Review and Maintenance -**Create a maintenance cycle for all AI models. Regularly review and update models to ensure they remain fit for purpose and prevent vulnerabilities or obsolescence. **iv) Impact Assessments-** Conduct impact assessments to evaluate the potential consequences of AI system failures. Identify critical areas where resilience is crucial. v) **Robust Security Measures**:- Implement robust security practices to safeguard against attacks. Address vulnerabilities and protect against adversarial threats2. **vi) System Robustness Strategies:** Develop strategies to enhance system robustness. Consider factors like data quality, model interpretability, and adaptability.	▪ **AI Governance Policies:** Ensure alignment with corporate strategy, risk management, and ethical implications. ▪ **Explainability and Transparency:** Guidelines for understanding and explaining AI decisions. ▪ **Risk and Compliance Monitoring:** Continuously monitor and address evolving aspects. ▪ **Resilience Assessment Tools Documentation to** analyze digital documents for fraud-resilient decision-making ▪ **Data Pipelines:** ▪ **AI Workload Documentation**

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
	5.1.4 AI controllability- AI controllability is the characteristic of an AI system whose functioning can be intervened by an external agen	i) **Ethical AI Design Principles**: ▪ Embed ethical considerations into AI development. ▪ Follow guidelines that prioritize fairness, privacy, and safety. ▪ Consider societal impact and unintended consequences. ii) **AI Alignment Strategies -** Create ways to ensure AI systems understand and follow human values. ▪ Align AI objectives with societal goals. ▪ Develop mechanisms for value preservation during AI training and decision-making. iii) **Transparent and Explainable -** Make AI systems transparent by revealing their inner workings. Understand how the model arrives at its predictions. Explain decisions to build trust and control. ▪ Use techniques like SHAP values, LIME, or attention mechanisms. iv) **Robust Testing and Validation**: ▪ Rigorously test AI systems under various scenarios. ▪ Validate their behavior against expected outcomes. ▪ Detect anomalies or unexpected behavior early. v) **Continuous Monitoring and Oversight**: ▪ Regularly monitor AI performance in production. ▪ Intervene proactively if issues arise. ▪ Ensure ongoing human involvement and regulatory action ▪ Determine who is offered what control over whose AI systems where multiple stakeholders are involved. ▪ Domain experts given the opportunity to provide feedback to not only re-assess the level of trust of the system but also to improve the operation of the system.	▪ **maintain records of the ethical considerations** integrated into the AI development process. These records may include documented discussions, decisions, and trade-offs related to fairness, privacy, and safety. ▪ **Guidelines Prioritizing Fairness, Privacy, and Safety:** These guidelines should explicitly address fairness, privacy protection, and safety measures. ▪ **Alignment with Societal Goals:** Organizations should document how AI objectives align with broader societal goals. This alignment ensures that AI systems contribute positively to societal well-being**.** ▪ **Value Preservation Mechanisms:** Records should capture the mechanisms implemented to preserve human values during ▪ **AI training and decision-making.** This includes documenting value alignment techniques and feedback loops. ▪ **Transparency Techniques:** Maintain documentation on the transparency techniques applied to AI models. This includes recording the use of methods like SHAP (SHapley Additive exPlanations) values, LIME (Local Interpretable Model-agnostic Explanations), or attention mechanisms. ▪ **Model Explanation Process:** Detailed records should explain how the model arrives at predictions. This

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
			documentation builds trust and allows for better control over AI systems. ▪ **Test Scenarios and Expected Outcomes:** records of the test scenarios used during AI system validation. These records should outline the expected behavior and outcomes under various conditions. ▪ **Anomaly Detection and Early Intervention:** Documenting the process of detecting anomalies or unexpected behavior during testing helps ensure robustness. Early intervention strategies should also be recorded. ▪ **Performance Monitoring Records:** maintain logs of AI system performance in production. These records help track deviations from expected behavior. ▪ **Human Involvement and Regulatory Action:** Document the roles of humans in monitoring and intervening when issues arise. Regulatory compliance efforts should also be recorded. ▪ **Stakeholder Control and Domain Expert Feedback:** Keep records of decisions regarding control over AI systems among stakeholders. Domain experts' feedback and assessments of trust levels should be documented
	5.1.5 AI explainability - AI explainability is the characteristic of an AI	i) **Model Selection and Simplicity:** ▪ Choose interpretable models whenever possible. Linear regression, decision trees, and rule-based models are more transparent than complex neural networks.	▪ **Policy Documents and Standards**: ○ **Explainability Policies**: Commissioned white papers, guidelines, and bills impact AI explanation

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
	system which can express important factors influencing a decision, prediction or recommendation in a way that humans can understand.	▪ Simplicity aids explainability. Avoid overfitting and excessive model complexity. ii) **Feature Importance:** ▪ Compute feature importance scores. Techniques like SHAP (SHapley Additive exPlanations) or feature importance from decision trees reveal which features influence predictions the most. ▪ Present these scores to users, highlighting the key factors driving decisions. iii) **Local Explanations:** ▪ Explain individual predictions. Techniques like LIME (Local Interpretable Model-agnostic Explanations) generate local explanations for specific instances. ▪ Show how input features contribute to a particular output. iv) **Global Explanations:** ▪ Provide an overview of model behavior. Aggregate feature importance scores across the entire dataset. ▪ Visualize global patterns and relationships. v) **Attention Mechanisms:** ▪ For neural networks, use attention mechanisms. These highlight relevant input features during prediction. ▪ Explain which parts of the input the model focused on. vi) **Rule-Based Systems:** ▪ Create rule-based decision systems. These are transparent and easy to understand. ▪ Define rules explicitly (e.g., "If feature A > threshold B, then predict class C"). vii) **Documentation and Reporting:** ▪ Maintain detailed documentation. Describe the model architecture, training process, and hyperparameters. ▪ Include explanations of preprocessing steps and any domain-specific considerations. viii) **User-Friendly Interfaces:** ▪ Design interfaces that display explanations to end-users.	practices. These documents outline requirements and expectations for explainability[1]. ○ **Standardization Documents**: Standards provide guidance on achieving explainability objectives. They address stakeholders' needs, including academia, industry, policymakers, and end-users[2]. ▪ **System-Level Documentation**: ○ **Full System View**: Document the entire AI system, including architecture, components, and interactions. This view helps estimate risks and ensures transparency. ○ **Provenance Documentation**: Record the lineage of data, models, and decisions. Provenance ensures traceability and reproducibility. ▪ **Model-Specific Documentation**: ○ **Model Architecture**: Describe the chosen model, its layers, and connections. ○ **Training Process**: Document hyperparameters, optimization techniques, and training data. ○ **Feature Importance**: Record feature importance scores (e.g., SHAP values) to explain predictions[5]. ○ **Local Explanations**: Explain individual predictions using techniques like LIME. ○ **Global Explanations**: Provide an overview of model behavior across the dataset. ▪ **User-Friendly Interfaces**:

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		▪ Use visualizations, natural language descriptions, or interactive tools. ix) **Feedback Loops:** ▪ Allow users to provide feedback on model predictions. ▪ Use this feedback to improve the model and address any discrepancies. x) **Ethical Considerations:** ▪ Explain how fairness, bias, and privacy were considered during model development. ▪ Document any trade-offs made to balance competing objectives.	○ **Explanatory Interfaces**: Design interfaces that display explanations to end-users. Use visualizations and natural language descriptions. ○ **Feedback Mechanisms**: Allow users to provide feedback on model predictions. ▪ **Ethical Considerations**: ○ **Fairness and Bias**: Document how fairness and bias were addressed during model development. ○ **Privacy Protection**: Explain how privacy concerns were considered. ▪ **Accountability and Responsibility**: ○ **Explanation Providers**: Allocate accountability to those responsible for providing explanations. ○ **Decision-Makers**: Document who makes decisions based on AI outputs.
	5.1.6 AI predictability - AI predictability is the characteristic of an AI system that enables reliable assumptions by stakeholders of its behaviour and the output.	i) **Clear Documentation and Communication:** ▪ Document Model Behavior: Describe the AI system's behavior, including its objectives, assumptions, and limitations. ▪ User-Friendly Explanations: Communicate with stakeholders in plain language. Explain how the AI arrives at decisions. ii) **Model Explainability Techniques:** ▪ Use techniques like SHAP values, LIME, or attention mechanisms to explain feature importance and prediction rationale.	○ **Model Behavior Document**: Detailed description of the AI system's behavior, including its objectives, assumptions, and limitations. ○ **User-Friendly Explanations**: Records of communication strategies used to explain the AI's decision-making process to stakeholders. ○ **Feature Importance Scores:** Documentation of feature importance (e.g., SHAP values) for each model. ○ **Local Explanations:** Records of individual prediction explanations (e.g., LIME results).

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		▪ Provide insights into which factors influence the AI's output. iii) **Risk and Safety Metrics:** ▪ Define and track risk metrics related to model performance. Assess the impact of incorrect predictions. ▪ Monitor safety metrics to ensure the AI system adheres to safety constraints. iv) **Stress Testing and Robustness Evaluation:** ▪ Conduct stress tests by subjecting the AI system to extreme conditions. Identify vulnerabilities and edge cases. ▪ Evaluate the AI's robustness across various scenarios and data distributions. v) **Traceability and Accountability:** ▪ Maintain a traceable record of model development, training data, and decision-making processes. ▪ Establish accountability by documenting who is responsible for the AI system. vi) **Risk Management Approach:** ▪ Implement a risk management strategy. Identify potential risks and develop mitigation plans. ▪ Regularly assess and update risk profiles. vii) **Transparency Reports:** ▪ Publish regular reports detailing the AI system's performance, updates, and any incidents. ▪ Include information on predictability and how the system aligns with stakeholder expectations. viii) **User Feedback and Iterative Improvement:** ▪ Gather feedback from users regarding the AI's behavior and predictions.	○ **Global Explanations:** Documentation of overall model behavior. ○ **Risk Metrics:** Regularly updated logs of risk-related metrics (e.g., false positives, false negatives). ○ **Safety Metrics:** Documentation of safety thresholds and adherence to safety constraints. ○ **Stress Test Results**: Detailed logs of stress tests, including extreme scenarios and edge cases. ○ **Robustness Assessment**: Documentation of robustness evaluations across different scenarios and data distributions. ○ **Model Development Timeline**: A traceable record of model development, including changes, updates, and versions. ○ **Decision Logs**: Documentation of key decisions made during model development and deployment. ○ **Risk Assessment Reports**: Regularly updated risk assessments, including identified risks and mitigation strategies. ○ **Risk Mitigation Plans**: Detailed plans for addressing potential risks. ○ **Transparency Reports**: Regularly published reports detailing AI system performance, updates, and incidents. ○ **Predictability Information**: Documentation on how the system aligns with stakeholder expectations.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		▪ Use this feedback to fine-tune the model and enhance predictability**.**	○ **User Feedback Logs**: Detailed records of user feedback regarding AI behavior, explanations, and satisfaction. ○ **Model Tuning History**: Documentation of model adjustments based on user feedback.
	5.1.7 AI transparency - AI transparency enables the stakeholders to be informed of the purpose of the AI system, how it was developed and deployed. This involves communicating information such as goals, limitations, definitions, assumptions, algorithms, data sources and collection, security, privacy and confidentiality protection and level of automation.	i) **Publish Information on Capabilities and Limitations:** ▪ Organizations should openly share details about what their AI systems can and cannot do. This includes both technical capabilities and practical limitations. ▪ Transparency reports or documentation should provide clear insights into the system's boundaries. ii) **Develop and Implement Reliable Content Detection Methods:** ▪ For audio-visual content generated by AI, consider watermarking or other techniques to identify synthetic content. ▪ Make these methods freely available to the public to enhance trust and accountability. iii) **Publish Training Data Description and Risk Mitigation Measures:** ▪ Describe the types of training data used to develop the AI system. Include information on data sources, diversity, and potential biases. ▪ Explain risk mitigation strategies employed during model development (e.g., fairness checks, bias reduction).	▪ **Clear Documentation:** ○ **Purpose and Goals**: Document the intended purpose of the AI system. Explain its objectives and how it aligns with organizational goals. ○ **Development Process**: Record the steps taken during development, including model selection, data preprocessing, and training. ○ **Deployment Details**: Document how the AI system is deployed, maintained, and updated. ▪ **Algorithm Descriptions:** ○ **Algorithms Used**: Clearly describe the algorithms employed. Explain their functioning and assumptions. ○ **Limitations**: Document algorithmic limitations, including scenarios where the model may fail or produce inaccurate results. ▪ **Data Sources and Collection:** ○ **Data Provenance**: Maintain records of data sources. Describe how data was collected, cleaned, and transformed.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		iv) **Clear Identification of AI Systems Mistaken for Humans:** ▪ Clearly label AI systems that interact with users as non-human. This prevents confusion and sets **appropriate expectations.** ▪ Prominently display disclaimers indicating that the system is an AI. v) **Assess Users' Satisfaction with Explanations:** ▪ Regularly collect feedback from users regarding the quality and comprehensibility of explanations provided by the AI. ▪ Use this feedback to improve the clarity and effectiveness of explanations. vi) **Reveal User's Mental Model of an AI System:** ▪ Understand how users perceive the AI system. Conduct surveys or interviews to uncover their mental models. ▪ Adjust communication strategies based on these insights. vii) **Assess User's Curiosity or Need for Explanations:** ▪ Gauge user curiosity about AI behavior and decision-making. Some users may seek detailed explanations, while others may prefer simplicity. ▪ Tailor explanations accordingly**.** viii) **Evaluate User's Trust and Reliance on the AI:** ▪ Assess whether users trust the AI system appropriately. Overreliance or blind trust can lead to unintended consequences. ▪ Monitor trust levels over time and address any issues.	○ **Bias and Fairness**: Document efforts to address bias and fairness issues in the data. ▪ **Model Explanations:** ○ **Feature Importance:** Explain which features influence predictions the most (e.g., SHAP values). ○ **Local Explanations**: Provide individual prediction explanations (e.g., LIME). ○ **Global Explanations:** Describe overall model behavior**.** ▪ **Assumptions and Definitions:** ○ **Assumptions Made:** Document any assumptions about the problem domain, user behavior, or data distribution. ○ **Key Definitions:** Clarify technical terms and concepts used in the AI system. ▪ **Security and Privacy:** ○ **Security Measures:** Detail security protocols to protect against unauthorized access or attacks. ○ **Privacy Protection:** Explain how user data is handled, anonymized, and secured. ▪ **Level of Automation:** ○ **Human-AI Interaction:** Specify the degree of automation. **Document when human intervention is required.** ○ **Decision Thresholds:** Describe decision thresholds and their impact on system behavior. ▪ **User-Friendly Communication:**

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		ix) **Assess Human-XAI Work System Performance:** ▪ Evaluate the overall performance of the human-AI collaboration. Consider factors like efficiency, accuracy, and user satisfaction. ▪ Continuously optimize the collaboration to achieve better outcomes.	○ **User Interfaces:** Design interfaces that convey transparency information to end-users. ○ **Explanatory Text:** Use natural language to explain system behavior and limitations. ▪ **Audit Trails and Logs:** ○ **Activity Logs:** Maintain logs of system activities, predictions, and user interactions. ○ **Model Updates:** Document model updates and version history. ▪ **Stakeholder Engagement:** ○ **Feedback Channels:** Establish channels for stakeholders to provide feedback. ○ **Regular Reporting:** Share transparency reports periodically.
	5.1.8 AI verification and validation - AI verification is the confirmation that an AI system was built right and fulfils specified requirements. AI validation is the confirmation with objective evidence that the requirements for a specific intended use of the AI application have been	i) **AI Verification:** ▪ **Test Item Quality Assessment**: Provide information about the quality of the test item (the AI system) based on how extensively it has been tested. ▪ **Residual Risk Evaluation**: Assess any remaining risks after testing. Identify areas where further testing or risk mitigation is needed. ▪ **Defect Detection:** Verify that defects (bugs, errors, inconsistencies) are identified and addressed before the AI system's release. ii) **AI Validation:**	▪ **Requirements Allocated to ML Component Management:** ○ **Requirements Document**: Detailed record of all requirements specific to the Machine Learning (ML) component. ○ **Traceability Matrix**: Mapping between requirements and ML components. ○ **Test Plans and Test Cases**: Documentation of how each requirement will be tested. ○ **Model Behavior Explanation**: Explanation of how the model behavior aligns with requirements.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
	fulfilled.	▪ **Objective Evidence:** Gather evidence that the AI system fulfills its intended purpose. This evidence can include test results, performance metrics, and user feedback. ▪ **Risk Mitigation:** Mitigate risks related to poor product quality. Address any gaps or issues identified during validation. ▪ **Stakeholder Satisfaction**: Validate that stakeholders' needs and expectations are met by the AI system. iii) **Accepted Software and Hardware Practices:** ▪ The AI system should adhere to established practices for software and hardware development. While AI components introduce unique challenges, they can still follow modified versions of these practices. ▪ **Unit Testing:** Test individual AI components (e.g., neural network layers, algorithms) to ensure correctness. ▪ **Functional Testing**: Validate that the AI system's functions work as expected. ▪ **Integration Testing:** Verify interactions between AI components. ▪ **Regression Testing:** Ensure that changes do not introduce new defects. ▪ **Performance Testing**: Assess system performance under different conditions. iv) **AI-Specific Validation Techniques:** ▪ **Empirical Testing**: Validate the AI system's behavior through real-world observations and experiments. ▪ **Intelligence Comparison:** Compare the AI's decision-making to human intelligence or established benchmarks.	▪ **Defect Detection and Risk Mitigation**: ○ **Defect Reports**: Detailed logs of defects identified during testing. ○ **Risk Assessment Reports**: Documentation of identified risks and mitigation strategies. ○ **Risk Mitigation Plans**: Plans for addressing potential risks. ▪ **Model Training and Robustness Evaluation**: ○ **Model Training Logs**: Detailed records of the training process, including hyperparameters and data used. ○ **Robustness Assessment Results**: Documentation of robustness evaluations across different scenarios. ○ **Model Performance Metrics**: Metrics related to accuracy, precision, recall, etc. ▪ **User Feedback and Iterative Improvement**: ○ **User Feedback Logs**: Detailed records of user feedback regarding AI behavior, explanations, and satisfaction. ○ **Model Tuning History**: Documentation of model adjustments based on user feedback. ▪ **AI-Specific Validation Techniques** (for pneumonia detection example): ○ **Validation Reports**: Detailed reports on how the pneumonia detector meets its intended purpose. ○ **Comparison to Human Intelligence**: Documentation of how the AI system performs relative to human experts.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		▪ Testing in Simulated Environments: Validate AI behavior in controlled simulations. ▪ **Field Trials**: Conduct real-world trials to assess performance and user satisfaction. ▪ **Comparison to Human Intelligence**: Evaluate how the AI system performs relative to human experts.	○ **Simulation Environment Logs**: Records of testing in simulated environments.
	5.1.9 AI bias and fairness A biased AI system can behave unfairly to humans (or certain subgroups). Fairness is a human perception and is based on personal and societal norms and beliefs. Unfair behaviour of AI systems can have negative, even harmful and devastative, impact on individuals or groups.	i) **Identify Bias Considerations** ▪ Explicitly document legal and ethical requirements related to bias. ▪ Include details on how bias considerations were factored into system requirements. • **Thresholds**: Set appropriate thresholds for acceptable bias levels. • **Stakeholder Involvement**: Document discussions with stakeholders regarding bias considerations. ii) **Provenance and Data Source Analysis** ▪ **Data Provenance Logs**: Maintain records of data sources, their origin, and any transformations applied. Include information on potential biases in the data. iii) Risk Assessment: Evaluate risks associated with data completeness and potential biases.	▪ **System Requirements Documentation** ○ legal and ethical requirements related to bias. ○ bias considerations were factored into system requirements. ○ thresholds set for acceptable bias levels ▪ **Data Provenance Logs** ○ Records of data sources, their origin, and any transformations applied ○ risks associated with data completeness and potential biases ○ data preprocessing steps to address bias. ▪ **Bias Testing Plans and Results:** ○ Detailed plans for testing bias in the AI system ○ Results of bias detection. ○ Fairness evaluation reports ▪ **Model Training Logs:**

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		iv) **Data Collection Process Review**: Document the review process for data collection and annotation. ▪ Maintain records of data sources, their origin, and any transformations applied. ▪ Include information on potential biases in the data. v) **Model Training Logs:** ▪ Document techniques used during model training to detect and mitigate bias. ▪ Record any adjustments made to algorithms to address bias. vi) **Bias Testing Plans and Results:** ▪ Create detailed plans for testing bias in the AI system. ▪ Document the results of bias testing, including any identified issues. vii) **Operational Review Logs:** ▪ Regularly review AI system behavior in real-world contexts. ▪ Record any bias-related issues encountered during operational reviews. viii) **User Feedback Logs:** ▪ Gather feedback from users regarding bias perceptions.	○ Records of techniques used during model training to detect and mitigate bias. ○ adjustments made to algorithms to address bias. ○ information on fairness metrics tracked during model development. ▪ **Operational Review Logs** ○ Review reports ○ Records of any bias-related issues encountered during operational reviews. ○ corrective actions taken ▪ **User Feedback Logs** ○ feedback from users regarding bias perceptions ○ user-reported bias incidents. ○ details on how user feedback influenced model adjustments. ▪ **External Toolkits and Resources:** ○ If you've used external toolkits (e.g., IBM AI Fairness 360, NIST resources), reference them in your documentation.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		▪ Document any user-reported bias incidents.	
5.2 Risks and risk management	**5.2.1 Risk management framework and processing –** The purpose of the Risk Management process is to identify, analyse, treat and monitor the risks continually. The Risk Management process is a continual process for systematically addressing risk throughout the lifecycle of an AI system product or service.	i) **AI Security Risk Assessment Framework:** ▪ **Risk Assessment** - AI risks should be identified, quantified or qualitatively described and prioritized against risk criteria and objectives relevant to the organization. ▪ **AI Asset Identification** - identify assets related to the design and use of AI that fall within the scope of the risk management process. ▪ **Controls** - identify controls relevant to either the development or use of AI, or both. Controls should be identified during the risk management activities and documented. ▪ **Identification of consequences** - As part of AI risk assessment, identify risk sources, events or outcomes that can lead to risks. Allso identify any consequences to the organization itself, to individuals, communities, groups and societies. ii) **Risk Analysis** ▪ **Assessing consequences –** when assessing consequences identified in the risk assessment, distinguish between a business impact assessment, an impact assessment for individuals and a societal impact assessment.	○ **Risk Assessment** ○ Detailed risk assessment reports. ○ Quantitative or qualitative descriptions of AI risks. ○ Prioritization against risk criteria and organizational objectives. ○ Risk registers or matrices. ○ Risk scoring or ranking documentation **AI Asset Identification**: ○ List of AI-related assets falling within the risk management scope. ○ Description of each asset's relevance to risk assessment. ○ Asset inventory. ○ Asset categorization. ○ **Controls** ○ Identification of controls relevant to AI development or use. ○ Documentation of control implementation. ○ Control descriptions. ○ Evidence of control effectiveness.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		▪ **Assessment of likelihood -** Where applicable, assess the likelihood of occurrence of events and outcomes causing risks. iii) **Risk treatment** ▪ Risk treatment options defined should be designed to reduce negative consequences of risks to an acceptable level, and to increase the likelihood that positive outcomes can be achieved iv) **Monitoring and review** ▪ continuous evaluation to ensure that the risk management framework remains effective. ▪ Regularly assess risk criteria, analysis, and treatment. ▪ Adapt to changes in external factors or organizational objectives. ▪ Involve stakeholders for holistic input. v) **Recording and reporting** ▪ establish, record, and maintain a system for the collection and verification of information on the product or similar products from the implementation and post-implementation phases. ▪ collect and review publicly available information on similar systems on the market. ▪ This information should then be assessed for possible	○ **Risk Analysis:** ○ Business impact assessment documentation. ○ Impact assessment for individuals and societal impact assessment. ○ Impact assessment reports. ○ Differentiated consequences analysis. ○ Assessment of Likelihood: ○ Likelihood assessment methods. ○ Probability estimates for risk events. ○ Likelihood assessments. ○ Probability distributions. ○ **Risk Treatment:** ○ Defined risk treatment options. ○ Strategies to reduce negative consequences and enhance positive outcomes. ○ Risk treatment plans. ○ Evidence of risk mitigation actions. ○ **Monitoring and Review:** ○ Continuous evaluation plans. ○ Criteria for assessing risk effectiveness. ○ Regular review reports. ○ Adaptation documentation.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		relevance on the trustworthiness of the AI system. In particular, the evaluation should assess whether previously undetected risks exist or previously assessed risks are no longer acceptable vi) Further guidance is available in Annex B, Kenya standard KS ISO/IEC 23894 and KS ISO 31000 .	○ Stakeholder involvement records. ○ **Recording and Reporting:** ○ System for collecting and verifying information on AI products. ○ Post-implementation phase reporting procedures. ○ Verification logs. ○ Post-implementation reports. ○ **Additional records** ○ a description and identification of the system that has been analysed; ○ methodology applied; ○ a description of the intended use of the AI system; ○ the identity of the person(s) and organization that carried out the risk assessment; ○ the terms of reference and date of the risk assessment; ○ the release status of the risk assessment; ○ — if and to what degree objectives have been met
5.3 Ethics and societal concerns -the presence, nature, extent and severity of an ethical concern with	**5.3.1 Ethical framework -** An AI ethical framework can be built on existing ethical frameworks such as virtual ethics,	i) **Accountability** ▪ Accountability occurs when an organization accepts responsibility for the impact of its actions on stakeholders, society, the economy and the environment.	1. **Accountability**: ○ **Documentation and Records**: ▪ **Accountability Framework**: Document how accountability is embedded in AI development and deployment.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
an AI system and application often depends upon the particular socio-political, economic, physical context of its development, implementation, audience or use. Further guidance is available in ISO/IEC ISO/IEC TR 24368:2022	utilitarianism, deontology and others Organizations contemplating the development and use of AI in responsible ways can consider adoption of various AI principles.	▪ Accountability provides necessary constraints to help limit potential negative outcomes and establish realistic and actionable risk governance for the organization. Combined, they help to define how to prioritize responsibilities. Some aspects that are covered by this theme are: o working with stakeholders to assess the potential impact of a system early on in the design; o — validating that stakeholder needs have actually been met; — o verifying that an AI system is working as intended; o ensuring the traceability of data and algorithms throughout the whole AI value chain; o enabling a third-party audit and acting on its findings; o providing ways to challenge AI decisions; o remedying erroneous or harmful AI decisions when challenge or appeal is not possible ii) **Safety and security** ▪ AI systems should be designed to be secure and resilient. This includes protecting against cyber-attacks and other security threats and their behavior in response to the range of tasks or situations to which they are likely to be exposed is understood. ▪ In addition to common IT security threats applicable to most systems (e.g. software bugs, hardware backdoors, data security breaches), certain AI systems, such as machine learning systems, can be vulnerable to specialized or targeted security threats. Such threats include the following: o data poisoning that results in a malfunctioning AI system;	▪ **Audit Trails**: Maintain records of decision-making processes, model training, and system behavior. ▪ **Responsibility Assignment**: Document roles and responsibilities of individuals involved in AI projects. ▪ **Safety Assessments**: Document safety considerations during AI design. ▪ **Security Protocols**: Record security measures implemented to protect AI systems. ▪ **Incident Response Plans**: Document procedures for handling security incidents. ▪ **Transparency and Explainability**: o **Model Documentation**: Detailed descriptions of AI models. o **Explainability Reports**: Document how model decisions are explained. o **Algorithmic Impact Assessments**: Record assessments of AI impact on stakeholders. ▪ **Fairness and Non-Discrimination**: o **Fairness Metrics**: Document fairness evaluations. o **Bias Mitigation Strategies**: Record steps taken to address bias. o **Non-Discrimination Policies**: Document policies against discriminatory outcomes. ▪ **Human Control of Technology**: o **Human Oversight Plans**: Document how humans maintain control over AI systems. o **Decision Points**: Record where human intervention is required.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		○ adversarial attacks that abuse a benign AI system; and ○ model stealing iii) **Fairness and non-discrimination** ▪ ensure that AI works well for people across different social groups, notably for those who have been deprived of social, political or economic power in their local, national and international contexts. ▪ These social groups differ across contexts and include but are not limited to those that require protection from discrimination based on sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation iv) **Transparency and explainability** ▪ ensure that people understand when they are interacting with an AI system, how it is making its decisions, and how it was designed and tested to ensure that it works as intended. ▪ The principle focuses on making sure an organization is transparent in its purposes and processes, whereas AI-specific principles focus on making sure that an AI system is understandable in how it works. v) **Ensuring human control of technology, over AI-Infused Systems** ▪ design AI systems and applications that enable human operators	○ **Human-AI Interaction Guidelines**: Document principles for human-AI collaboration. ▪ **Professional Responsibility**: ○ **Ethical Charters**: Record ethical guidelines for AI development. ○ **Professional Codes of Conduct**: Document adherence to professional standards. ○ **Training Records**: Maintain records of AI professionals' training on ethical practices. ▪ **Promotion of Human Values**: ○ **Value Alignment Statements**: Document how AI systems align with human values. ○ **Stakeholder Feedback**: Record input from users and communities. ○ **Value-Driven Objectives**: Document objectives related to societal well-being. ▪ **International Human Rights**: ○ **Human Rights Impact Assessments**: Record assessments of AI impact on human rights. ○ **Adherence to International Standards**: Document alignment with global human rights norms. ○ **Human Rights Due Diligence**: Maintain records of due diligence efforts. ▪ **Respect for International Norms of Behavior**:

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		to review or authorize automated decisions; ▪ allow for the ability to opt in or opt out of automated decisions; ▪ critically evaluate how and when to delegate decisions to AI systems and applications, and how such systems and applications can transfer control to a human in a manner that is meaningful and intelligible. vi) **Professional responsibility,** ▪ The theme of professional responsibility aims to ensure that professionals who design, develop or deploy AI systems and applications or AI-based products or systems, recognize their unique position to exert influence on people, society and the future of AI - especially since policies, norms and principles often lag behind new and emerging technologies. . vii) **Promotion of human values,** ▪ Ensure that AI is deployed and utilized in a way that maximizes benefit to society, promote humanity's wellbeing and encourage human flourishing ▪ Particular applications of AI that aim to promote human values include (but are not limited to) ○ improving health and healthcare; ○ improving living situations; ○ improving working conditions;	○ **Cross-Cultural Considerations**: Document how AI behavior aligns with global norms. ○ **Cultural Context Assessments**: Record assessments of cultural implications. ○ **Norms Adherence Reports**: Document adherence to international norms. ▪ **Community Involvement and Development**: ○ **Stakeholder Engagement Plans**: Document community involvement strategies. ○ **Community Feedback Logs**: Record input from affected communities. ○ **Community Impact Assessments**: Assess AI impact on local communities. ▪ **Respect for the Rule of Law**: ○ **Legal Compliance Reports**: Document adherence to legal frameworks. ○ **Legal Assessments**: Record assessments of legal risks. ○ **Legal Opinion Letters**: Maintain legal opinions related to AI compliance. ▪ **Sustainable Environment**: ○ **Environmental Impact Assessments**: Document AI impact on the environment. ○ **Energy Efficiency Measures**: Record efforts to minimize resource usage.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		○ environmental and sustainability efforts. viii) **International human rights,** ▪ . UN Guiding Principles on Business and Human Right], are fundamental moral principles to which a person is inherently entitled, simply by virtue of being human. They can serve as a guiding framework for directing corporate responsibility around AI systems and applications with the benefit of international acceptance as a more mature framework for assessments of policy and technology International ix) **Respect for international norms of behaviour, community involvement and development,** ▪ complying with laws and regulations even when they are not enforced in that jurisdiction; ▪ abiding by all legal obligations throughout the whole AI value chain and periodically reviewing compliance of the stakeholder's activities and relationships; ▪ ensuring the purposes for which AI is developed and used to be lawful and specified. x) **Respect for the rule of law,** ▪ The rule of law demands, inter alia, that even powerful organizations and systems comply with the law. ▪ Compliance with legal requirements, including recourse to judicial redress as appropriate against decisions rendered by AI systems and applications, is an established aspect of information and communication technology, data governance and risk management.	○ **Sustainability Reports**: Assess AI's contribution to environmental sustainability. ▪ **Labor Practices**: ○ **Fair Labor Policies**: Document fair treatment of workers. ○ **Worker Rights Compliance**: Record adherence to labor laws. ○ **Worker Well-Being Initiatives**: Document efforts to support workers.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		Following the rule of law in each jurisdiction in which it operates can include: ○ complying with laws and regulations even when they are not enforced in that jurisdiction; ○ abiding by all legal obligations throughout the whole AI value chain and periodically reviewing compliance of the stakeholder's activities and relationships; ○ ensuring the purposes for which AI is developed and used to be lawful and specified xi) **Environmental sustainability** ▪ One sustainability dilemma challenging data- and computing-intensive technologies, such as AI, is the ever-increasing need for energy resources as large data sets and algorithms require consumption of even greater amounts of processing power. This increased need is occurring even as global sustainable development goals call for energy efficiency and lowering non-renewable consumption. ▪ Hence, the importance to offer transparent information to stakeholders about energy consumption, climate change and the mitigation of adverse impacts across the AI-based service value chain, in order to enable stakeholders to make sustainable decisions. ▪ An organization can also utilise AI systems and applications to foster sustainability and manage environmental impacts and climate change, through a life cycle approach aimed at reducing waste, reusing products and components, and recycling materials. Examples include: ○ energy grid optimisation, ○ precision agriculture,	

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		○ sustainable supply chains, ○ climate monitoring, and ○ environmental disaster prediction xii) **labour practices** ▪ The International Labour Organization, and the UN tripartite agency, brings together governments, employers and workers to set labour standards, develop policies and devise programmes that are adopted by consensus, to promote decent work. ▪ Potential considerations regarding the role AI plays in labour relations include: ○ ensuring that the rules regarding employment and employment relationships are understood and that humans are involved in the type of decisions that require effective human oversight and empathy (for example the use of AI in managing workers, including gig-economy workers, avoiding discrimination between workers, preventing disproportionate and undue surveillance at work, particularly in remote work, protecting worker privacy, eliminating all forms of forced or compulsory labour and the effective abolition of child labour); ○ assuring fair remuneration, working conditions and health and safety, workers protection and other concerns are addressed, (for example, crowd sourced and outsourced workers preparing training data or content moderators exposed to AI-mediated social media content); ○ issues of human development and training, especially in a setting where the introduction of AI eliminates work roles or changes their nature in a major fashion (for example, retraining); ○ anticipating the consequences of the introduction of AI and	

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		reskilling of the workforce; ○ assurances that respect for human life and human dignity are maintained and that AI and big data systems do not negatively affect human agency, liberty and dignity; ○ providing for rules on businesses’ and developers’ liability; — making AI the subject of social dialogue and collective bargaining according to the rules and practices in place in each organization	
	5.3.2 Societal concerns - Common ethical concerns relate to the means of collecting, processing and disclosing of personal data, conceivably with biased opinions, that feed opaque machine learning decision-making algorithms which are not explainable.	i) **Privacy** ▪ Privacy aims to ensure that AI systems and applications are developed and implemented with natural persons’ right to privacy in mind, as well as deceased persons’ (through the executor of their estate or nominee as applicable). ▪ Right to privacy has become one of the most prominent themes in AI development, due in large part, to the Data Protection Regulation in Kenya. Common dimensions of privacy include: ○ limiting data sourced, collected, used or disclosed to that which is necessary for accomplishing the intended purposes and tasks; ○ communication of the purpose of the processing of personal identifiable information and any sharing of it; ○ consent: transparency on the data held on a natural or a deceased person, natural or deceased persons’ data not to be collected or used without their knowledge or permission; ○ control over the use of data: natural or deceased persons’	▪ **Privacy Impact Assessments (PIAs)**: ○ Conduct PIAs before deploying AI systems. ○ Document the assessment process, findings, and mitigation strategies. ○ PIA reports with identified privacy risks and recommended actions. ▪ **Data Minimization and Purpose Specification**: ○ Clearly define the purpose of data collection. ○ Document the minimum necessary data required for AI training. ○ Purpose statements. ○ Data minimization policies. ▪ **Consent Records**: ○ Document user consent for data processing. ○ Specify the scope of consent (e.g., training, profiling). ○ Consent forms or mechanisms.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
		control of the use of its personal identifiable ○ information; ○ natural or deceased persons' degree of influence over how and why their information is used; ○ ability to restrict data processing: natural or deceased persons' power to have data restricted from collection or use in connection with AI technology; ○ **rectification**: enable natural or deceased persons to modify information if it is incorrect; ○ **erasure**: enable natural or deceased persons to remove personal data from an AI system and application; ○ enabling natural or deceased persons to view personal data used by an AI system and application; ○ **privacy by design**: integrating considerations of data privacy into the development of AI systems and applications and throughout the overall life cycle of data use ○ **dispute resolution**: offer mechanisms for resolving disputes in relation to these features.	○ Timestamps of consent. ▪ **Privacy Policies and Notices**: ○ Maintain clear and concise privacy policies. ○ Document how personal data is handled. ○ Privacy policy documents. ○ Updates and revisions. ▪ **Data Retention Policies**: ○ Define data retention periods. ○ Document the rationale for retention. ○ Retention schedules. ○ Data deletion logs. ▪ **Anonymization and Pseudonymization Techniques**: ○ Describe methods used to protect privacy. ○ Document anonymization processes. ○ Anonymization guidelines. ○ Evidence of pseudonymization. ▪ **Privacy by Design Documentation**: ○ Describe privacy features integrated during AI system design. ○ Document privacy-enhancing technologies. ○ Privacy by design reports. ○ Privacy-enhancing tool usage.

Characteristic	Sub-Characteristic	Measures and Activities	Output and Documentation
			▪ **Incident Response Plans**: ○ Develop plans for handling privacy breaches. ○ Document roles, procedures, and communication channels. ○ Incident response playbooks. ○ Incident logs.
	5.3.3 Legal requirements and issues	▪ AI technology is new and the legal requirements associated with its development, deployment and use are not yet widely defined. Some regions have instituted legal requirements governing certain aspects of AI technology and applications (e.g. facial recognition for law enforcement), and a wide range of proposals has been made and debated. Currently there are no coordinated and cohesive legal requirements at the domain, regional, national or international levels concerning AI technology.	▪ **Data protection Act** ▪ **ICT policy**

6 Conformance

6.1 To claim conformance to this code, implementors of this code are required to commit to adopting the identified measures.

6.2 The code identifies measures that should be applied in advance of binding regulation pursuant to existing Artificial Intelligence and Data regulations by all firms developing or managing the operations of a generative AI system with general-purpose capabilities, that are made widely available for use, and which are therefore subject to a wider range of potentially harmful or inappropriate use.

6.3 Organizations developing and managing the operations of these systems both have important and complementary roles. Developers and managers need to share relevant information to ensure that adverse impacts can be addressed by the appropriate firm.

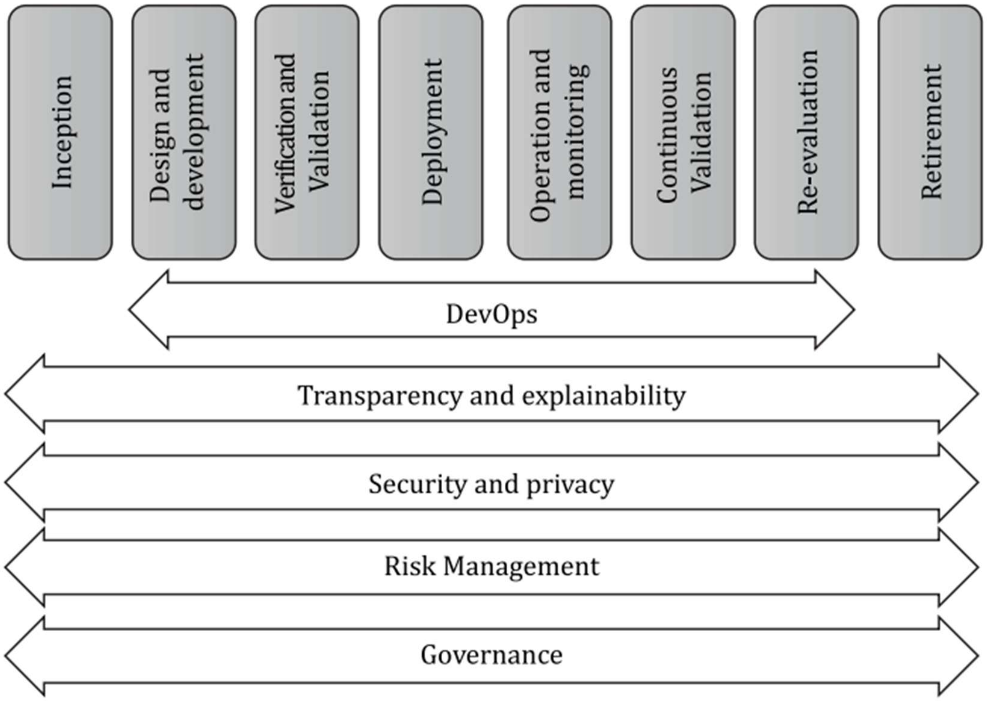


Figure 1 — An Example of AI Life Cycle Processes

Annex A
(normative)

Stakeholder roles and responsibilities

A.1 General

This clause provides recommendations for the stakeholders to recognize their roles and responsibilities as well as be made aware of opportunities in making, using or responding to the impact of the AI application.

A.1.1 AI producer perspective

The AI producer should at least address the following considerations:

— Who are the AI customers and AI users?

— Who are the AI developers? Are they qualified and skilled employees or contractors?

— Who are the AI application providers and their relationship with the AI producer?

— Who are the stakeholders in each stage of the AI system life cycle?

What is the AI system and its capabilities? What algorithm is the AI model based on?

— What are the AI characteristics of the AI application?

— What data are used to create the AI model? What is the source of these data? Who are the data providers and their partners?

— What are the trustworthiness and risk concerns of the AI application? What is being done to assess and mitigate these concerns? Is a risk management system in place for the organization?

— What are the ethics, societal concerns, security, confidentiality, privacy and other legal requirement considerations in producing and deploying the AI application? How are they being addressed?

— What is the technological ecosystem for the accessible deployment of the AI application?

— What is the overall quality of the AI system?

— How is the AI application built, applied and updated? How is the AI model trained or programmed? How robust is the AI mode? When (in which stage of AI system life cycle) the model building, application and updates will be reviewed? Where is the model built, applied and updated, on site or using a cloud service? When (in which stage of AI system life cycle) should the AI producer be involved to reassess AI characteristics in context?

— Where is the AI application to be deployed, on-premise or as a cloud service? Where will the AI application be developed? Where are the AI developers located? Where are the data sources located? Where, in terms of, geographical location will the AI application be deployed?

— Why is the AI application being developed into a product or service? What is the potential value for the AI producer and AI customer? What are the opportunities and courses of action?

A.1.2 Data provider perspective

The data provider should at least address the following considerations:

— Who is the AI producer? Employer, partner or customer?

— What data are being collected and what is the source? How are the data collected, stored, processed, provisioned and fed into the AI model (for machine learning applications) Is a data management system employed?

— What is the domain, geographical and other providence of the data being collected? What are the applicable boundary conditions of the AI model developed from these data?

— What are the sources and nature restrictions for gathering the required training data?

— How is the quality of the collected data measured and validated? What are the trustworthiness and bias concerns of the data? What is being done to assess and mitigate these concerns?

— How are data being collected, validated and used to update the AI model during the operation and

maintenance stage? How are collected data secured, protected and used appropriately in compliance

with internal policies and data sovereignty requirements?

— When (in which stages of AI system life cycle) the data availability and quality need to be reassessed?

— Where is the source location of the data? Where are the data to be processed, on-premise or using a cloud

service? In which geographic location?

— Why specific data are needed in the context of the AI application?

A.1.3 AI developer perspective

The AI developer should at least address the following considerations:

— Who is the AI user, data provider and AI producer?

— What is the relationship between the AI developer and AI producer? Employee or contractor?

— What are the qualifications and skills required of the AI developers?

— What AI model is employed, trained or programmed? How is the AI model being designed, developed,

validated and verified into the functional characteristics of the AI system? What processes are involved?

— What are the technological and ecosystem requirements needed to deploy the AI system as an accessible AI application?

— What are the algorithms used for data processing? What are the criteria for data quality? What are the criteria for output quality? What are criteria for validation and verification? What are the criteria for model update?

— How are data pre-processed? How is the quality of data determined? How is the algorithm selection done? How are the model requirements adapted?

— When (in which stage of AI system life cycle) are the context and requirements assessed?

— Where the AI application can be deployed, locally or as a cloud service?

— Why is the AI application being developed into a product or service? Why the specific model is used?

A.1.4 AI application provider perspective

The AI application provider should at least address the following considerations:

— Who are the AI customers and AI users and how do they employ the AI application?

— What is the relationship between the AI producer and the AI application provider? Employer or partner?

— What are the AI characteristics of the application? What are its capabilities, capacity and throughput as well as constraints and limitations?

— What are the technological and ecosystem requirements for the AI customers and AI users to access and use the AI application? What are the failure recovery provisions?

— What are the operational analytics of the AI application and how are they monitored?

— What are the impacts of the AI application on its customers, users and community?

— How is the AI application built, applied and updated? How is the AI model trained or programmed? How robust is the AI model, When (in which stages of AI system life cycle) the model building, application and updates are to be reviewed? Where is the model built, applied and updated, on site or as a cloud services? When (in which stages of AI system life cycle) should the producer be involved to reassess AI characteristics in context?

— How are risks managed in the deployment of the AI application?

— When (in which stage of AI system life cycle) are the context and requirements assessed?

— Are there any applicable boundaries for the recommended, acceptable or responsible use of the AI application? Are these part of the legal requirements in the software license?

— Where is the AI application being deployed? What legal requirements apply to the functional and non functional characteristics of the AI application's domain? Who are the regulators?

— Why is the AI application being developed into a product or service?

A.2 Use perspective

A.2.1 General

The stakeholders with the use perspective are those AI customers and AI users that employ the AI application to augment their decision-making.

A.2.2 AI customer and AI user perspective

The AI customer and AI user should at least address the following considerations:

— What is the relationship between the AI application provider and AI customer or AI user?

— What are the AI customers' and AI users' (and as community members) considerations in using the AI application? What are some of the governance implications involved in organizations where the AI application is employed?

— What data are collected in using the AI application and how are they being used (for machine learning applications, see Reference? What are the data governance policies in place? Are the data being fed back into the AI model for continuous learning and improvement?

— What are the trustworthiness and risk considerations of the AI application being used? What is being done to assess and mitigate these concerns?

— What are the transparency and explainability aspects of the AI application supplied by the AI provider?

— What are the ethical and societal concerns in using the AI application? How are they addressed?

— What decision-making will be augmented by the AI application? What is the level of automation? Who is going to evaluate the effectiveness of the AI application and what metrics are being used?

— How do the AI customers and AI users access the output of the AI application to augment their decision making? How are the performance and effectiveness of the AI application being measured?

— When (in which stage of AI system life cycle) are the context and requirements assessed? Where is the AI application deployed and accessed? What are the legal requirements for deployment?

— Why is the AI application being employed? What are the potential values in employing the AI application?

A.3 Impact perspective

A.3.1 General

The community in which the AI application is deployed and its consumers in it can be impacted by its use. Examples include the use of AI applications in surveillance, loan application, delivery of health care, information dissemination in social media. The deployment of an AI application can be impacted by the regulator who is an authority in the locality and has jurisdiction governing the use of AI technology based on legal requirements promulgated by policy makers.

A.3.2 Community perspective

The community in which the AI application is deployed should at least address the following considerations:

— Who are the consumers? What are their particular concerns as a member of the community?

— What data are collected in using the AI application and how are they being used? What are the privacy concerns?

— How is the community and consumers in it being impacted by the employment of the AI application? How is this impact being measured, how often and by whom? What are the community's recourses for adverse impacts?

— When (in which stage of AI system life cycle) the AI customer feedback or requirements are to be assessed and reassessed?

A.3.3 Regulator and policy maker perspective

Regulators and policy makers should at least address the following considerations:

— Who are the consumers? What are their particular concerns as a member of the community?

— What is the mechanism through which legal requirements are made for the deployment of the AI application (e.g. top-down or bottom-up)? How is the AI application being used and how does the employment impact the community? Who is the responsible party (e.g. AI provider, AI customer, AI user)?

— When (in which stage of AI system life cycle) are the legal requirements assessed or reassessed?

— Where is the AI application being deployed? What are the applicable legal requirements? How is the deployment going to be monitored for compliance? Who is the responding party for a violation?

— Why is the AI application being employed? What are the potential values in employing the AI application? What are potential, positive or adverse impacts on the community

Annex B

(normative)

AI Quality Assurance Model

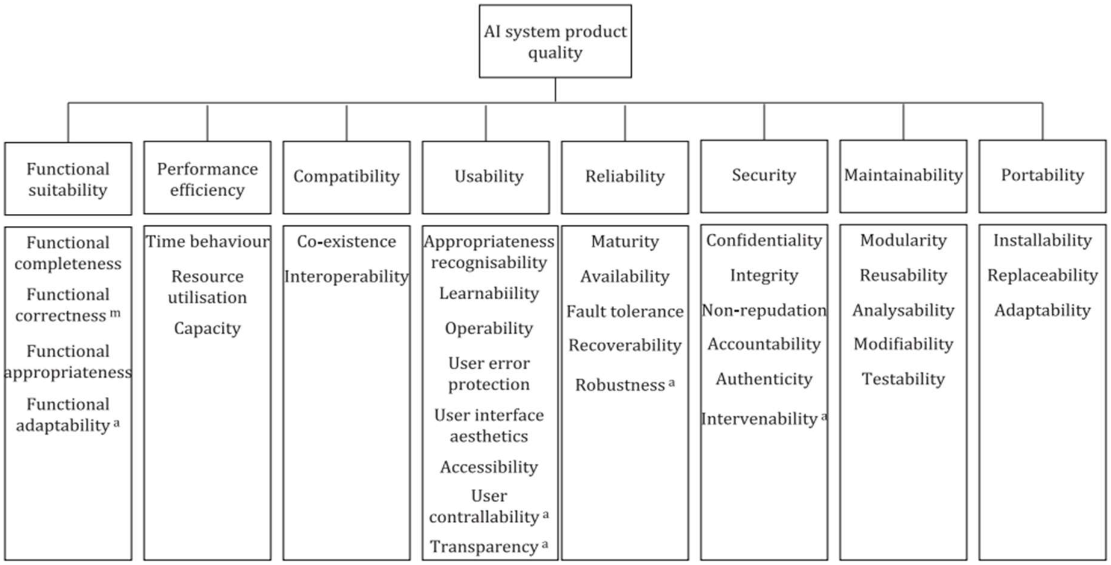
B.1 Quality characteristics

The quality characteristics of the quality model of AI systems are useful to elicit and identify quality requirements of non-functional requirements, which are often implicit stakeholder needs. Refer to Kenya Standard KS ISO/IEC 25059.

B.2 Product quality model

B.2.1General

An AI system product quality model is detailed in Figure 1. The model is based on a modified version of a general system model provided in Kenya Standard KS ISO/IEC 25010. New and modified sub-characteristics are identified using a lettered footnote. Some of the sub-characteristics have different meanings or contexts as compared to the KS ISO/IEC 25010 model. The modifications, additions and differences are described in this clause. The unmodified original characteristics are part of the AI system product model and shall be interpreted in accordance with Kenya Standards KS ISO/IEC 25010



[a] New sub-characteristics.

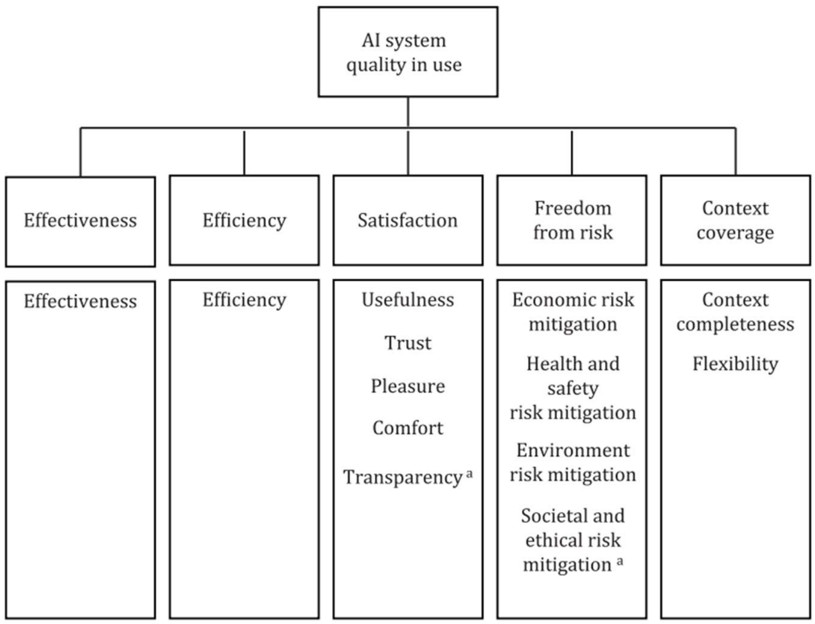
[m] Modified sub-characteristics.

Figure B.1 — AI System Product Quality Model

B.3 Quality in use model

B.3.1 General

An AI system quality in use model is detailed in Figure 2. The model is based on a modified version of a general quality in use model provided in KS ISO/IEC 25010. New sub-characteristics are identified using a lettered footnote. Some of the sub-characteristics have different meanings or contexts as compared to the Kenya Standard KS ISO/IEC 25010 model. The additions and differences are described in this clause. The unmodified characteristics are part of the quality in use model and shall be interpreted as defined in KS ISO/IEC 25010



New sub-characteristics.

Figure B.2— AI Quality in Use Model

Annex C

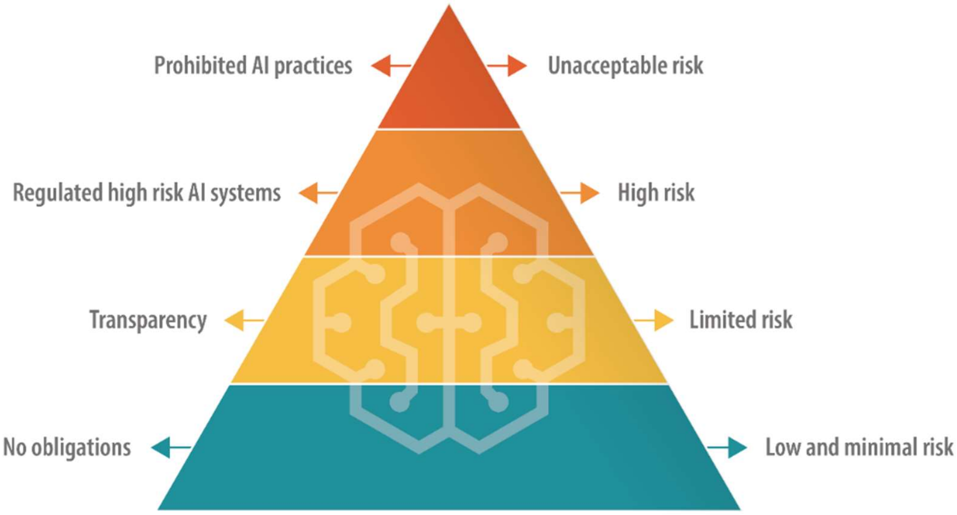
Risk Management Framework

(Normative)

B.1 AI risk assessment

AI risks should not be considered in isolation. Different AI actors have different responsibilities and awareness depending on their roles in the lifecycle. For example, organizations

developing an AI system often will not have information about how the system may be used. AI risk management should be integrated and incorporated into broader enterprise risk management strategies and processes. Treating AI risks along with other critical risks, such as cybersecurity and privacy, will yield a more integrated outcome and organizational efficiencies.



Data source: European Commission.

Annex C
(Informative)

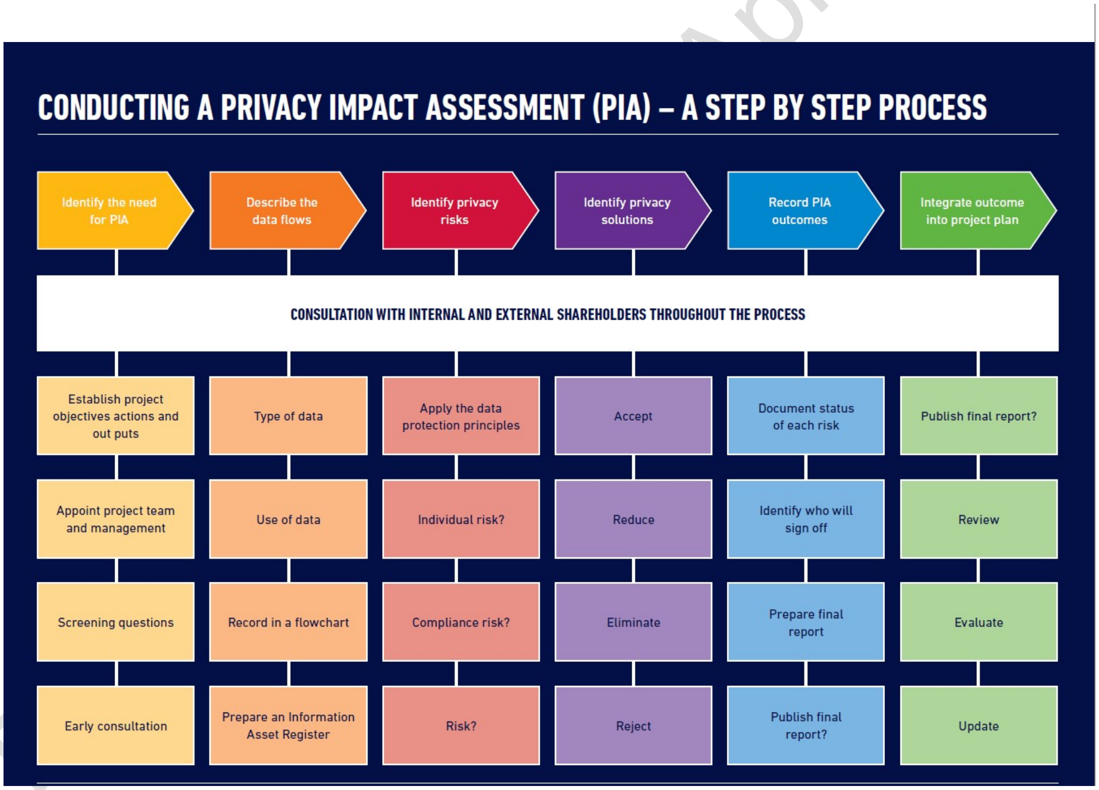
Privacy Risk Impact Assessment

C.1 Privacy values such as anonymity, confidentiality, and control generally should guide choices for AI system design, development, and deployment. Privacy-related risks may influence security, bias, and transparency and come with trade-offs with these other characteristics.

C.2 Like safety and security, specific technical features of an AI system may promote or reduce privacy. AI systems can also present new risks to privacy by allowing inference to identify individuals or previously private information about individuals

C.3 Privacy risk management is a cross-organizational set of processes that helps organizations to understand how their systems, products, and services may create problems for individuals and how to develop effective solutions to manage such risks.

Figure C2- Data Privacy Impact Assessment



CONDUCTING A PRIVACY IMPACT ASSESSMENT (PIA) – A STEP BY STEP PROCESS

CONSULTATION WITH INTERNAL AND EXTERNAL SHAREHOLDERS THROUGHOUT THE PROCESS

Identify the need for PIA	Describe the data flows	Identify privacy risks	Identify privacy solutions	Record PIA outcomes	Integrate outcome into project plan
Establish project objectives actions and out puts	Type of data	Apply the data protection principles	Accept	Document status of each risk	Publish final report?
Appoint project team and management	Use of data	Individual risk?	Reduce	Identify who will sign off	Review
Screening questions	Record in a flowchart	Compliance risk?	Eliminate	Prepare final report	Evaluate
Early consultation	Prepare an Information Asset Register	Risk?	Reject	Publish final report?	Update