

OFFICE OF THE DATA PROTECTION COMMISSIONER

Guidance Note on Artificial Intelligence

November 2026

CONTENTS

DEFINITION OF TERMS 1

FOREWORD 5

OFFICE 6

INTRODUCTION 7

Background 7

Privacy and Data Protection Concerns in AI 7

Scope and Purpose of Guidance Note 8

Proportionate Application 9

LEGISLATIVE FRAMEWORK 10

The Constitution of Kenya, 2010: 10

The Data Protection Act, Cap 411C: 10

The Data Protection (General) Regulations, 2021 11

The Data Protection (Registration of Data Controllers and Data Processors) Regulations, 2021. 11

The Data Protection (Complaints Handling Procedure and Enforcement) Regulations, 2021..... 11

Other Applicable Laws 11

APPLICATION OF DATA PROTECTION PRINCIPLES TO AI SYSTEMS 12

Lawfulness, Fairness, and Transparency 12

Purpose Limitation 13

Data Minimisation 13

Accuracy 14

Storage Limitation 14

Integrity and Confidentiality 15

Accountability 16

LAWFUL BASIS FOR PROCESSING PERSONAL DATA IN AI SYSTEMS 17

Consent 17

Contract 17

Legal Obligation 18

Vital Interests 18

Public Interest or Exercise of Official Authority 18
Legitimate Interests 19
Lawful Basis for Training, Fine-Tuning and Evaluation 19
RIGHTS OF DATA SUBJECTS IN AI CONTEXTS **21**
Right to Be Informed 21
Right of Access 21
Right to Rectification 21
Right to Erasure 22
Right to Restriction 22
Right to Object 22
Right Not to Be Subject to Automated Decision-Making 22
Right to Data Portability 24
OBLIGATIONS OF DATA CONTROLLERS AND DATA PROCESSORS IN AI SYSTEMS **26**
Registration 26
Privacy by Design and Default 26
Data Protection Impact Assessment (DPIA) 26
Data Protection Officer (DPO) 27
Sensitive Personal Data and AI 27
Children and AI Systems 28
Notification and Communication of a Breach 29
Fairness, Non-Discrimination and Accessibility 29
Data Sharing and Third-Party AI Processors 30
Cross-Border Transfers of AI-Processed Data 31
DATA PROTECTION AND THE AI LIFECYCLE **33**
Stage 1: Problem Definition and AI System Design 34
Stage 2: Data Collection and Preparation 34
Stage 3: Model Training and Development 35
Stage 4: Testing, Validation, and Pre-Deployment Review 36
Stage 5: Deployment and Integration 37

Stage 6: Operation, Monitoring, and Maintenance 38

Stage 7: Decommissioning and Data Disposal 39

HIGH-RISK AI APPLICATIONS: SPECIFIC OBLIGATIONS 41

PUBLIC SECTOR ARTIFICIAL INTELLIGENCE AND DIGITAL PUBLIC INFRASTRUCTURE 44

PRE-PROCESSING CHECKLIST FOR AI SYSTEMS..... 46

AI GOVERNANCE FRAMEWORK 48

GENERATIVE AI 48

AGENTIC ARTIFICIAL INTELLIGENCE AND RECOMMENDER SYSTEMS 50

ENFORCEMENT AND COMPLAINTS..... 51

ANNEX 1: COMPLIANCE CHECKLIST FOR AI SYSTEMS..... 53

ANNEX 2: DPIA TRIGGER TABLE FOR AI USE CASES..... 55

DEFINITION OF TERMS

"Act" means the Data Protection Act, Cap. 411C.

"Agent ic AI" means an artificial intelligence system that can pursue a goal by planning and executing a sequence of actions with limited or no human intervention, including by invoking software tools, transacting, or otherwise acting on external systems on behalf of a user or an entity.

"AI Deployer" means an entity that uses an artificial intelligence system during its operations to process personal data or to produce outputs, recommendations, or decisions affecting data subjects.

"AI Developer" means an entity that designs, builds, trains, or substantially modifies an artificial intelligence system or artificial intelligence model, whether for its own use or for provision to others.

"AI lifecycle" means the successive and iterative phases through which an artificial intelligence system passes, including its design, planning, data collection and processing, development, training, testing, validation, deployment, operation, monitoring, modification and decommissioning.

"AI Provider" means an entity that makes an artificial intelligence system or artificial intelligence model available to others, directly or through an intermediary, including through artificial intelligence-as-a-service platforms.

"AI User" means a natural person who interacts with or operates an artificial intelligence system under the authority of an AI deployer.

"Algorithmic Transparency" means the disclosure of meaningful information about the logic, significance, and envisaged consequences of artificial intelligence-driven decision-making to data subjects and the public.

"Anonymisation" means the removal of personal identifiers from personal data so that the data subject is no longer identifiable.

"Artificial Intelligence (AI)" means a machine-based system that can, for a given set of objectives, make predictions, recommendations, or decisions influencing real or virtual environments, operating with varying levels of autonomy.

"Automated Decision-Making (ADM)" means the process of deciding solely by automated means, without any meaningful human involvement.

"Biometric Data" means personal data resulting from specific technical processing relating to the physical, physiological, or behavioural characteristics of a natural person which allows or confirms the unique identification of that person.

"Data Commissioner" means the person appointed pursuant to Section 6 of the Act.

"Data Controller" means a natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purpose and means of processing of personal data.

"Data Processor" means a natural or legal person, public authority, agency, or other body which processes personal data on behalf of the data controller.

"Data Protection Impact Assessment (DPIA)" means a process designed to identify and mitigate data protection risks arising from a processing activity, required under Section 31 of the Act.

"Data Protection Officer (DPO)" means the officer designated pursuant to Section 24 of the Act to advise on and monitor compliance with data protection obligations.

"Data Subject" means an identified or identifiable natural person who is the subject of personal data.

"Digital Public Infrastructure (DPI)" means shared digital systems that enable the delivery of public and private services at population scale, including digital identity systems, digital payment systems and data exchange platforms.

"Explainability" means the degree to which the internal mechanics of an artificial intelligence system can be understood and communicated in meaningful human terms to those affected by its outputs.

"**Foundation model**" means an artificial intelligence model that is trained on a broad range of data, generally at scale, and that displays significant generality such that it is capable of being used, with or without modification, to perform a wide range of distinct tasks, or of being integrated into a variety of downstream systems or applications.

"Foundation Model Provider" means an entity that develops or provides a large artificial intelligence model trained on broad datasets at scale, including a large language model, that is capable of being adapted to a wide range of downstream tasks.

"Generative AI" means artificial intelligence systems capable of generating new content, including text, images, audio, video, or code, based on training data.

"High-Risk AI Application" means an artificial intelligence system whose use is likely to result in significant risks to the fundamental rights, freedoms, or vital interests of data subjects, including in the areas of credit scoring, employment, healthcare, criminal justice, and children's services.

"Inferred Data" means personal data that is derived, predicted or otherwise produced by an artificial intelligence system about a data subject rather than obtained directly from that data subject, including risk scores, behavioural profiles and predicted attributes.

"Large Language Model (LLM)" means a type of artificial intelligence system trained on large volumes of text data, capable of generating, summarising, classifying, or translating natural language.

"Machine Learning (ML)" means a subset of artificial intelligence in which systems learn from data to improve their performance on a task without being explicitly programmed.

"Meaningful Human Review" means review of an automated decision by a natural person who has the competence, authority, information and time to assess the decision independently and to confirm, vary or reverse it. Review that is nominal, automatic, or exercised without the capacity to alter the outcome is not meaningful human review.

"Model Card" means a structured record documenting an artificial intelligence model's intended purpose, training data sources and characteristics, evaluation results including performance disaggregated by relevant groups, limitations and known risks.

"Office" means the Office of the Data Protection Commissioner established under Section 5 of the Act.

"Personal Data" means any information relating to an identified or identifiable natural person.

"Privacy by Design and by Default" means the integration of data protection principles and technical measures into the design of artificial intelligence systems and processing activities from the outset.

"Processing" means any operation or set of operations performed on personal data, whether by automated means, including collection, recording, organisation, storage, adaptation, retrieval, use, disclosure, combination, restriction, erasure, or destruction.

"Profiling" means any form of automated processing of personal data to evaluate certain personal aspects relating to a natural person, to analyse or predict aspects concerning performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location, or movements.

"Pseudonymisation" means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, and such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data is not attributed to an identified or identifiable natural person.

"Recommender System" means an artificial intelligence system that ranks, curates, personalises or otherwise determines the content, offers or options presented to a data subject.

"Red-Team Testing" means structured adversarial testing of an artificial intelligence system by persons acting independently of the development team to identify security, privacy, safety and fairness failures before and during deployment.

"Regulations" means all regulations made in accordance with Section 71 of the Act.

"Sensitive Personal Data" means data revealing a natural person's race, health status, ethnic or social origin, conscience, belief, genetic data, biometric data, property details, marital status, family details, sex, or sexual orientation.

"Synthetic Data" means data generated by a model or algorithm to reproduce the statistical properties of a source dataset without corresponding to any actual data subject, and which remains personal data where an individual can be re-identified from it or from its combination with other information.

"Training Data" means the dataset used to train an artificial intelligence or machine learning model, which may contain personal data.

FOREWORD

Artificial Intelligence is rapidly reshaping how individuals, organisations, and governments interact with information and make decisions. From credit scoring and healthcare diagnostics to predictive policing and online content curation, AI systems are processing vast volumes of personal data in ways that can profoundly affect the rights, freedoms, and opportunities of data subjects. This presents both enormous potential and significant risk for the privacy and dignity of Kenyans.

The right to privacy is guaranteed under Article 31 of the Constitution of Kenya, 2010 and is given specific effect by the Data Protection Act, Cap 411C. As AI technologies become embedded in public services and private enterprise alike, the Data Protection Act applies with full force to the collection, use, and automated processing of personal data by AI systems. The Office of the Data Protection Commissioner (Office) is mandated under the Act to regulate and enforce compliance with these obligations.

This Guidance Note on Artificial Intelligence has been developed to provide data controllers, data processors, developers, and deployers of AI systems with clear and practical direction on their obligations under the Data Protection Act and the attendant Regulations. It applies data protection principles to the distinctive characteristics and risks of AI processing, including algorithmic transparency, automated decision-making, profiling, and the use of sensitive personal data in machine learning models.

The Office recognises that AI innovation is essential to Kenya's Bottom-Up Economic Transformation Agenda (BETA) and the Digital Superhighway initiative. It is the firm view of this Office that robust data protection is not an obstacle to AI-driven development but its necessary foundation. Responsible AI begins with responsible data governance.

This Guidance Note sets out the principles and obligations that should guide all AI processing of personal data. Entities that deploy AI systems should regard this Note as a minimum standard and are encouraged to supplement it with sector-specific safeguards, technical controls, and governance frameworks.

I call upon all organisations operating AI systems in Kenya to engage with this Guidance Note and to work proactively with the Office to build an AI ecosystem that is lawful, fair, transparent, and accountable.



Immaculate Kassait, SC, MBS

Data Commissioner

11th November 2026

OFFICE

The Office of the Data Protection Commissioner (the Office) is a government agency established to protect the privacy and security of personal data in our increasing digital world. It is responsible for enforcing data protection laws and policies to safeguard the privacy, dignity, and fundamental rights of individuals. The Office is mandated to oversee the implementation and enforcement of the Data Protection Act, 2019 (the Act), which regulates the processing of personal data of persons located in Kenya by both private and public sector organisations.

The Office plays a vital role in ensuring that individuals have control over their personal data and that organisations respect their privacy rights. The Office's work involves monitoring and enforcing compliance with data protection regulations, investigating data breaches, and imposing sanctions on entities that violate data protection laws. In addition, the Office is responsible for raising public awareness about data protection issues and educating individuals and organisations on how to protect personal data. With the growing importance of data protection in our digital age, the Office is a critical institution in maintaining trust and confidence in our data-driven society.

The Office is uniquely positioned to facilitate both the government and private sector entities in achieving Government's strategic goals under the "Bottom-Up Economic Transformation Agenda" and, in particular, its digital superhighway initiative. As the digital landscape expands, the need for robust data protection mechanisms becomes paramount. The Office, with its mandate to oversee, regulate, and ensure lawful data processing, plays a pivotal role in this transformation.

Kenya remains at the cutting edge of digital transformation while maintaining stringent data protection standards. The Office serves as a key stakeholder and regulator in guiding the nation's digital superhighway journey by ensuring that as we advance technologically, the rights and privacy of individuals remain safeguarded.

INTRODUCTION

Background

AI is increasingly deployed across Kenya's public and private sectors. Financial institutions use AI-driven credit scoring and fraud detection. Hospitals and health insurers apply machine learning to diagnostic imaging and patient risk stratification. Law enforcement agencies explore predictive analytics. Employers use algorithmic tools to screen job candidates. Retailers and digital platforms deploy AI for personalised advertising and content recommendation. Government agencies process citizen data through AI-enabled service delivery and benefits administration systems.

This widespread adoption of AI is occurring against the backdrop of Kenya's ambitious digital transformation agenda and growing data infrastructure, including expanding mobile internet penetration, the proliferation of digital financial services, and increasing government digitisation. The volume of personal data being collected, aggregated, and fed into AI training pipelines and inference systems is growing exponentially.

AI systems by their nature introduce novel data protection challenges that existing sectoral guidance has not fully addressed: the opacity of algorithmic models, the aggregation of large training datasets, the generation of inferences and predictions about data subjects without their direct input, the risk of discriminatory outcomes from biased training data, and the reduced role of human decision-makers in consequential choices. These characteristics require specific data protection guidance calibrated to the AI context.

AI systems process personal data in ways that extend beyond data directly provided by individuals. In addition to collected data, these systems generate inferred data such as behavioural profiles, risk scores, and predictions about individuals. Although not explicitly supplied by the data subject, such inferences relate to identifiable people and therefore constitute personal data under Section 2 of the Act. Their use raises important transparency and fairness obligations under Sections 25 and 29, particularly where individuals may be unaware that such conclusions are being drawn.

AI systems also produce outputs, including classifications, recommendations, and automated decisions that may significantly affect individuals. Where these outputs relate to identifiable persons, they fall within the scope of personal data and are subject to the protections of the Act. Organisations should therefore ensure compliance not only at the point of data collection, but across the full AI processing chain, including the use of outputs and the safeguards required for automated decision-making under Section 35.

Privacy and Data Protection Concerns in AI

The deployment of AI systems raises a distinct and significant set of privacy and data protection concerns:

- **Opacity and lack of explainability**: Many AI models, particularly deep learning systems, operate as ***"black boxes",*** making it difficult to explain how a particular output or decision was reached. This fundamentally challenges the transparency principle and the right of data subjects not to be subjected to solely automated decisions without explanation.

- **Discriminatory profiling and bias:** AI systems trained on historical data can perpetuate and amplify existing patterns of discrimination, resulting in profiling that disproportionately disadvantages individuals based on race, gender, ethnicity, disability, or economic status, characteristics that intersect with sensitive personal data categories under the Act.
- **Excessive data collection for training:** The training of large AI models often requires vast datasets of personal data, creating pressure to collect and retain more data than would be justified under the data minimisation and storage limitation principles.
- **Re-identification risk:** Ostensibly anonymised datasets used for AI training can be re-identified through linkage attacks or inference from model outputs, undermining the effectiveness of anonymisation as a de-risking measure.
- **Cross-border data flows:** AI services are frequently provided by offshore processors, and training data may be transferred to and processed in foreign jurisdictions without adequate data protection frameworks, contrary to the Act's transfer restrictions.
- **Children and vulnerable groups:** AI systems may process children's data or be deployed in contexts affecting vulnerable individuals, in education, child welfare, or social protection without appropriate safeguards calibrated to heightened risk.
- **Generative AI and synthetic data:** Large language models and generative AI systems may reproduce or recombine personal data from their training sets in their outputs and may generate realistic but false content about real individuals.
- **Inferred data and AI-generated outputs:** AI systems may derive new information about individuals, including behavioural profiles, risk scores, and predictions, and generate outputs such as classifications or decisions that significantly affect individuals. These inferences and outputs may remain opaque to data subjects despite influencing outcomes.
- **Emerging AI risks (bias, drift, and security):** AI systems may introduce risks such as proxy discrimination, feedback loops that reinforce disadvantage, degradation in accuracy over time (model drift), and exposure to attacks including data extraction, model inversion, or adversarial manipulation.

Scope and Purpose of Guidance Note

The purpose of this Guidance Note is to provide data controllers, data processors, AI developers, and deployers with an understanding of their obligations under the Act and the attendant Regulations when developing, deploying, or procuring AI systems that process personal data of persons located in Kenya.

This Guidance Note applies to all entities, whether public or private, that:

- develop AI systems that are trained on personal data,
- deploy AI systems that process personal data in producing outputs, recommendations, or decisions,
- develop, provide, fine-tune or host foundation models or generative artificial intelligence systems that process personal data,
- operate recommender systems that rank, curate or personalise content, offers or options using personal data,

- deploy agentic artificial intelligence systems capable of taking actions on external systems on behalf of a data subject or an entity,
- procure AI-as-a-service products or platforms that involve the processing of personal data, or
- use AI systems in automated decision-making that affects the rights or interests of data subjects.

This Guidance Note applies throughout the lifecycle of an artificial intelligence system, from problem definition and data collection through training, testing, deployment, operation and decommissioning. It applies to personal data used as an input to an artificial intelligence system, to personal data contained in or recoverable from a trained model, and to inferred data and other outputs generated by the system that relate to an identified or identifiable natural person. It applies to public and private entities alike.

This Guidance Note should be regarded as a minimum standard which can be supplemented by additional measures. Entities operating in regulated sectors should additionally comply with sector-specific data protection guidance issued by this Office and applicable sector regulators.

This Guidance Note considers:

a) The Constitution of Kenya, 2010;
b) The Data Protection Act, Cap 411C;
c) The Data Protection (General) Regulations, 2021;
d) The Data Protection (Registration of Data Controllers and Data Processors) Regulations, 2021;
e) The Data Protection (Complaints Handling Procedure and Enforcement) Regulations, 2021;
f) Other applicable laws and regulations; and
g) International best practices and comparative AI governance frameworks.

Proportionate Application

The obligations in this Guidance Note apply proportionately, having regard to the nature, scope, context and purpose of the AI system; the type and volume of personal data processed; the size and resources of the data handler; and the likelihood and severity of harm to data subjects. Before processing begins, each data handler should classify and document the risk presented by the AI system. Small data handlers may use simplified policies, records and governance arrangements where these remain effective and appropriate to the identified risk.

The size of a data handler does not, however, determine the risk of its processing. An AI system should be treated as high-risk where it processes sensitive personal data or data relating to children or other vulnerable persons; involves systematic monitoring, profiling or biometric identification; materially influences access to a service, benefit or opportunity; operates at scale or in a publicly accessible place; supports critical or digital public infrastructure; or combines datasets in a manner that materially increases what may be inferred about an individual. In such cases, the data handler should implement safeguards proportionate to the actual risk, irrespective of its size or resources.

LEGISLATIVE FRAMEWORK

The development, deployment, and use of AI systems in Kenya is governed by a layered legislative framework. The following instruments are of primary relevance:

The Constitution of Kenya, 2010:

Article 31 guarantees every individual the right to privacy, including in respect of information relating to their family or private affairs.

Article 27 guarantees equality and non-discrimination, providing a constitutional basis for challenging AI systems that produce discriminatory profiling or decisions.

Article 46 guarantees consumers' rights to safe and transparent goods and services, and provides for redress, including compensation, where loss or injury arises from defective or harmful outcomes.

Article 35 guarantees access to one's personal data held by another person, requiring disclosure where rights are affected, and allowing individuals to correct or delete inaccurate data

The Data Protection Act, Cap 411C:

The Act is the primary statutory instrument governing AI processing of personal data. Key provisions include:

- Designation of a Data Protection Officer (DPO) for entities whose core activities involve large-scale systematic monitoring or the large-scale processing of sensitive personal data, including many high-risk AI deployments.
- Application of the core data protection principles to all AI processing activities.
- Protection and facilitation of data subject rights.
- Transparency obligations, including informing data subjects about automated processing and profiling.
- Lawful bases for AI processing, including consent, contract, legal obligation, vital interests, public task, and legitimate interests.
- Requirement to conduct a Data Protection Impact Assessment (DPIA) for high-risk AI processing, including automated decision-making and large-scale profiling.
- Enhanced protection of children's personal data, including parental or guardian consent and age-appropriate safeguards for AI systems accessible to children.
- Protection against solely automated decision-making that significantly affects a data subject, including the right to meaningful human review and an explanation of the logic underlying the decision.
- Right to object to the processing of personal data, including processing for AI profiling.
- Right to data portability, where AI systems rely on personal data provided by the data subject.
- Restrictions on cross-border transfers of personal data, including transfers involving AI cloud services and offshore AI providers.

- Enforcement powers of the Data Commissioner, including issuing enforcement notices, imposing administrative penalties, and ordering the cessation of unlawful AI-related processing.

The Data Protection (General) Regulations, 2021

These Regulations elaborate the principles and obligations under the Act. They set out the content requirements for privacy notices, the conditions for valid consent, the requirements for Data Protection Impact Assessments, and the obligations of data processors engaged through contractual arrangements.

All these obligations apply directly to AI processing activities.

The Data Protection (Registration of Data Controllers and Data Processors) Regulations, 2021

All entities deploying or operating AI systems that process personal data of data subjects in Kenya should register with the Office as data controllers or data processors, subject to applicable thresholds and exemptions.

The Data Protection (Complaints Handling Procedure and Enforcement) Regulations, 2021

These Regulations set out the procedure for lodging, admitting, and determining complaints, the conduct of investigations, and the enforcement mechanisms available to the Office, including enforcement and penalty notices. They apply fully to AI systems that process personal data, including complaint handling, investigations, and enforcement of data protection obligations arising from AI-driven processing and decisions.

Other Applicable Laws

Entities deploying artificial intelligence should comply with this Guidance Note and all applicable legal, regulatory and policy requirements, including but not limited to those concerning accessibility, non-discrimination, cybersecurity, communications, consumer protection, access to information, employment and workplace monitoring. They should also comply with requirements specific to the sector in which the system is used, such as health, payment systems, financial services or education. Where more than one regulatory framework applies to the processing of personal data, the entity should comply with each and apply the requirement that affords greater protection to data subjects.

APPLICATION OF DATA PROTECTION PRINCIPLES TO AI SYSTEMS

All processing of personal data by or in connection with AI systems should comply with the data protection principles set out in Section 25 of the Act. The following sub-sections address the specific application of each principle to AI contexts.

Lawfulness, Fairness, and Transparency

AI systems should process personal data lawfully, which requires identifying and relying on a valid lawful basis under Section 30 of the Act.

Fairness requires that AI systems do not produce outcomes that arbitrarily disadvantage data subjects or use their data in ways that would be unexpected or harmful.

Transparency requires that the use of AI, the categories of personal data involved, and the logic of automated processing be disclosed to data subjects in accessible, plain-language terms.

Entities deploying AI systems should ensure that their privacy notices specifically disclose:

- that personal data is processed by automated or AI-assisted means;
- the categories of data used as inputs;
- the material factors that influence the output or decision and their relative significance, expressed in plain language;
- the known limitations of the system, including its accuracy range and the categories of case in which it is less reliable;
- whether and at what point a natural person reviews the output, and how the data subject may request human review;
- that the data subject is interacting with an artificial intelligence system rather than a natural person, where that is the case;
- that content presented to the data subject has been generated or materially altered by an artificial intelligence system, including synthetic audio, image or video depicting a real person;
- the nature of the outputs, inferences, or decisions produced; and
- the significance and envisaged consequences of such processing for the data subject.

Privacy notices should be layered, concise and expressed in plain language.

Example

Jua Kali Credit, a digital lender, uses an AI-based credit scoring model that analyses mobile money transaction history and device usage patterns to determine loan eligibility. To comply with the lawfulness, fairness, and transparency principle, the company:

- identifies and documents a valid lawful basis before deployment (performance of the loan contract for application data, and a documented Legitimate Interests Assessment for fraud checks);
- ensures fairness by testing the model to confirm that it does not produce unjustified adverse outcomes for groups of applicants and that the data used is within applicants' reasonable expectations; and

- ensures transparency through a privacy notice that discloses the use of AI scoring, the categories of data used as inputs, the significance of the score for lending decisions, and the applicant's right to seek human review of an adverse automated decision under Section 35 of the Act.

Purpose Limitation

Personal data collected for a specific purpose should not subsequently be used to train AI models for a different, incompatible purpose. Entities should assess whether the use of existing datasets for AI training is compatible with the original collection purpose, considering the nature of the data, the reasonable expectations of data subjects, and the potential consequences for data subjects of further processing.

Where personal data is to be used to train AI models, entities should identify and document a specific, lawful purpose for that training activity. Repurposing production or operational data for AI training without fresh legal basis or compatibility assessment is not permitted.

Example

A public hospital collects patient records under a lawful basis of public interest for the delivery of healthcare services. The hospital cannot, without further assessment and legal authority, supply those records as training data to a commercial AI vendor developing a diagnostic tool for commercial licensing. The use of the data for commercial AI development is incompatible with the original public healthcare purpose.

Data Minimisation

AI systems should be designed and operated to use only personal data that is adequate, relevant, and necessary to achieve the defined purpose. Entities should resist the pressure to collect and process ever-larger datasets on the assumption that more data should produce better AI outcomes.

Privacy by design requires that data minimisation be addressed at the AI system design stage, not retrospectively.

Where aggregated, pseudonymised, or synthetic data can achieve the same training or operational purpose as full personal data, entities should prefer those less privacy-invasive approaches.

Entities should maintain a record of the provenance of each dataset used to train, fine-tune or evaluate an artificial intelligence system, identifying its source, the date of acquisition, the lawful basis relied upon, the categories of personal data it contains and any licence or contractual restriction on its use. Personal data obtained by scraping websites, social media or other publicly accessible sources remains personal data, and its use requires a lawful basis; public accessibility does not of itself make the processing lawful.

Example

A digital lender builds an AI model for credit scoring but collects additional data such as contact lists and social media activity that are not necessary for assessing creditworthiness. This violates the data minimisation principle.

The lender should restrict data collection to relevant financial and transactional information.

Accuracy

AI systems can amplify inaccuracies in training data and generate inaccurate inferences at scale. Entities should take reasonable steps to ensure that personal data used to train AI models is accurate and up-to-date and should implement mechanisms to detect and correct inaccurate outputs that affect data subjects.

Erroneous outputs, particularly in credit scoring, health diagnosis, or fraud detection can cause serious harm. Entities should have processes to rectify inaccurate AI-generated information about data subjects upon request under Section 40 of the Act.

Example

An insurance company uses an AI system trained on outdated data, resulting in incorrect risk assessments and higher premiums for some customers.

This undermines the accuracy principle. The company should regularly update data and allow correction of inaccurate information affecting decisions.

Example

An AI screening tool trained on outdated or incomplete CVs may wrongly infer that a candidate lacks a required qualification, misread an employment gap as a reliability risk, or rank applicants down because of errors in the data it was given. Because the tool processes hundreds of applications at once, a single inaccuracy can unfairly exclude many qualified candidates without any human noticing.

Employers using AI in recruitment should verify candidate data against source documents where possible and allow applicants to review and correct information that informed a rejection. Where a candidate shows that the system relied on inaccurate data, the employer should rectify or delete the erroneous record and reconsider the application.

Storage Limitation

Personal data used to train AI models should not be retained longer than is necessary for the training purpose. Entities should establish and enforce data retention policies that address AI training datasets specifically, including deletion or anonymisation of training data upon completion of the relevant training cycle. Model weights and parameters that encode personal information should also be addressed in retention and deletion policies.

Retention and deletion obligations extend to every copy of personal data associated with an artificial intelligence system, including training and validation datasets, feature stores, embeddings and other vector representations, prompt and output logs, caches, backups and copies held by processors and sub-processors. Entities should be able to demonstrate that deletion has been effected across all such copies and should record any copy that cannot be deleted, the reason, and the compensating controls applied.

Example

A telecommunications company retains personal data used to train an AI model even after the model has been deployed and the data is no longer needed.

This contravenes the storage limitation principle. The company should delete or anonymise the data once the training purpose is fulfilled.

Integrity and Confidentiality

AI systems should incorporate appropriate technical and organisational security measures to protect personal data against unauthorised access, loss, or destruction at all stages of the AI lifecycle, including data collection, preprocessing, training, inference, and storage of outputs. Entities should specifically address the security of:

- training datasets and the systems used to store and process them;
- model parameters and weights, which may be subject to model inversion or membership inference attacks capable of reconstructing training data;
- prompts, retrieval sources and system instructions, which may be manipulated through direct or indirect prompt injection;
- tools, plug-ins and external systems that an agentic artificial intelligence system is authorised to invoke;
- unapproved or unmanaged use of third-party artificial intelligence services by staff, commonly described as shadow artificial intelligence, through which personal data may be disclosed to third parties without authorisation;
- inference APIs and the personal data transmitted through them; and
- logs of AI-generated outputs that may contain or relate to personal data.

Entities should assess and address security risks that are specific to artificial intelligence, including training data poisoning, adversarial and evasion inputs, model inversion and membership inference, model and prompt extraction, unauthorised fine-tuning, and abuse of tool or agent permissions.

For high-risk artificial intelligence applications, entities should conduct red-team testing before deployment and at intervals proportionate to the risk, maintain tamper-evident logs of system inputs, outputs and actions taken, maintain an incident response playbook addressing artificial intelligence failures including erroneous and discriminatory outputs, and establish a channel through which security researchers and members of the public may report vulnerabilities and harmful outputs.

Example: A healthcare provider uses an AI system to analyse patient data, but weak access controls allow unauthorized access to sensitive records.

This breaches the integrity and confidentiality principle. The provider should implement strong access controls and security measures to protect the data.

Accountability

Entities deploying AI systems bear accountability for ensuring that the system complies with data protection requirements throughout its lifecycle. Accountability obligations in the AI context include:

- maintaining records of processing activities that document AI systems and their data inputs;
- conducting DPIAs before deploying high-risk AI applications;
- assigning documented accountability for each artificial intelligence system to a named senior officer;
- maintaining an inventory or register of artificial intelligence systems in use, recording the purpose, risk classification, data sources, model or service provider, human review arrangements and date of last assessment for each system, including artificial intelligence obtained as a feature of a wider software product;
- reporting periodically to the board or equivalent governing body on the artificial intelligence systems in use, their risk classification and any incidents arising;
- implementing governance structures that assign clear responsibility for AI data protection compliance; and
- monitoring deployed AI systems on an ongoing basis for drift, bias, or emerging privacy risks.

Example

A financial institution deploys an AI fraud detection system without documenting the data used, system design, or measures to ensure fairness.

This undermines accountability. The institution should maintain records, conduct assessments, and implement governance structures to demonstrate compliance

LAWFUL BASIS FOR PROCESSING PERSONAL DATA IN AI SYSTEMS

Every processing operation conducted by or in connection with an AI system should have a valid lawful basis under Section 30 of the Act. Entities should identify the lawful basis before processing commences and should document that identification. The following sub-sections address the application of each lawful basis to AI processing.

Consent

Consent, under the Act, should be freely given, specific, informed, and unambiguous.

In the AI context, consent provides a valid lawful basis only where it is obtained in relation to the specific AI processing and decision-making concerned. Before giving consent, the subject of data should be provided with clear information about:

- the use of AI;
- the categories of data processed by the AI system; and
- the nature of any automated decision-making or profiling.

Transparency obligations under Section 29 of the Act apply irrespective of the lawful basis relied upon; the provision of a privacy notice does not, of itself, constitute consent. Where consent is relied upon, entities should be able to demonstrate that it was freely given, specific, informed, and unambiguous in respect of the AI processing, including any automated decision-making or profiling.

Data subjects should be informed of their right to withdraw consent at any time, and withdrawal should not result in detriment.

Consent is not an appropriate lawful basis where there is a significant power imbalance between the entity and the data subject (for example, employer-employee AI monitoring) or where consent would need to be given as a condition of access to a service.

Example

Optional AI spending insights: A mobile banking app offers an optional AI feature that analyses transaction history to suggest a budget. Before enabling it, the bank explains that AI will use the customer's transaction categories and amounts to infer spending patterns, and that it will not make decisions about credit eligibility. The customer turns the feature on through a separate, unticked choice; receiving ordinary banking services does not depend on that choice. The bank records the consent and lets the customer withdraw it in the app without losing access to their account. Its general privacy notice informs customers about processing but does not serve as consent to this optional feature.

Contract

AI processing may be based on contractual necessity where the processing is objectively necessary for the performance of a contract with the data subject or to take steps at the request of the data subject prior to entering into a contract.

Example

A lender may rely on contractual necessity to process an applicant's financial data where assessing the loan application is necessary to take steps requested by the applicant before entering into a loan agreement. This may include reviewing information about the applicant's income, existing debts and repayment history to determine whether to offer a loan and on what terms. If the lender uses an AI system to assist with that assessment, it should limit the data supplied to the system to what is needed for the application.

Legal Obligation

Where a statutory or regulatory obligation requires an entity to conduct a processing activity, and that activity is performed using AI, the legal obligation lawful basis may be relied upon.

Example

A regulated financial institution uses an AI tool to flag unusual transactions for review as part of its monitoring obligations under the Proceeds of Crime and Anti-Money Laundering Act. It may rely on legal obligation for the personal data processing necessary to meet those obligations.

Vital Interests

AI systems in emergency health contexts or disaster response may process personal data based on vital interests where the processing is necessary to protect the life of a data subject or another natural person. This basis should be relied upon only where other bases are not available, and processing is genuinely necessary to protect life.

Example

Flood rescue: During a flood, a rescue team uses an AI system to analyse live drone footage and identify people stranded on rooftops. It processes their images and locations so responders can find them quickly when seeking consent is impracticable. The team limits access to the footage, uses it for the rescue operation and retains it only for as long as necessary or otherwise required by law.

Public Interest or Exercise of Official Authority

Public sector entities and entities performing public functions may rely on public interest or official authority as a lawful basis for AI processing that is necessary for the performance of those functions. Public bodies should nonetheless adhere to all other data protection principles and obligations, including transparency and DPIA requirements.

Example

Disease outbreak monitoring**:** A county public health department uses an AI system to analyse disease reports from health facilities and identify a possible outbreak. Where this processing is

necessary to carry out its legally assigned public health functions, it may rely on public interest or official authority.

Legitimate Interests

Private sector entities may rely on legitimate interests where the processing is necessary for the purposes of the legitimate interests pursued by the data controller or a third party, provided those interests are not overridden by the interests or fundamental rights of the data subject.

Entities relying on legitimate interests for AI processing should conduct and document a Legitimate Interests Assessment (LIA) that:

- identifies the legitimate interest;
- demonstrates necessity; and
- balances the interest against the data subject's rights, considering the intrusiveness of AI profiling and automated decision-making.

Legitimate interests require careful assessment and may not be appropriate for certain high-risk AI applications, such as those that profile data subjects at scale, involve sensitive personal data, or produce significant automated decisions affecting data subjects' legal rights or opportunities, depending on the nature of the processing and its impact on data subjects.

Example

Detecting fraudulent transactions: An online retailer uses an AI system to flag unusual purchase patterns for fraud review. Its legitimate interest is preventing fraud, and it documents why analysing limited transaction details is necessary for that purpose. In its assessment, the retailer considers the risk of incorrectly flagging a customer, limits the data used, and ensures that a person reviews a flagged transaction before an order is cancelled. It informs customers about the processing and does not use the fraud data to build unrelated marketing profiles.

Lawful Basis for Training, Fine-Tuning and Evaluation

The training, fine-tuning and evaluation of an artificial intelligence model are distinct processing activities, each of which requires its own lawful basis. An entity that collected personal data to provide a service may not rely on that collection alone as authority to use the data to train a model. It should identify a lawful basis for the training activity and assess whether the training purpose is compatible with the purpose for which the data was collected.

Where an entity relies on legitimate interests to train a model on personal data, the Legitimate Interests Assessment should address the reasonable expectations of the data subject, the availability of anonymised, pseudonymised or synthetic alternatives, the risk that the model may memorise and reproduce personal data, and the means by which a data subject may object to the use of their personal data for training.

Where an entity procures a foundation model or an artificial intelligence service from a third party, it should obtain and record information about the provenance of the data used to train that model

and the lawful basis relied upon by the provider and should assess whether reliance on the model is consistent with its own obligations under the Act.

Example

Fine-tuning a customer support chatbot: A company collects customer messages to answer support queries. It later proposes using those messages to fine-tune a chatbot and test its responses. Before doing so, it treats fine-tuning and evaluation as separate uses, assesses whether they are compatible with the original support purpose, and identifies an appropriate lawful basis for each. If it relies on legitimate interests, it documents customers' reasonable expectations, whether anonymised or synthetic examples would suffice, the risk of the model reproducing personal details, and how customers can object. If a third-party provider supplies the underlying model, the company also seeks information about its training data and provenance.

RIGHTS OF DATA SUBJECTS IN AI CONTEXTS

The Act provides data subjects with enforceable rights that apply with full force to AI systems processing their personal data. Entities deploying AI systems should implement technical and procedural mechanisms to give effect to each of the following rights as well as resolve the data subject requests within the timelines prescribed under the Act and the Data Protection (General) Regulations, 2021. Where a request is declined, the entity should notify the data subject of the refusal in writing, with reasons, within the prescribed period.

Right to Be Informed

Under Section 29 of the Act, data subjects should be informed at the point of data collection of the use of AI systems, the logic of automated processing, the significance of any automated decisions, and the envisaged consequences of the processing for the data subject.

Privacy notices should address AI processing in terms that are intelligible to a lay person.

Entities should not rely on generic or technical language that obscures the nature of AI decision-making.

Where an artificial intelligence system is used to make or materially inform a decision about a data subject, the entity should explain that use in its privacy notice in clear, plain language and in a form accessible to people with limited digital access or skills. The entity should also provide the data subject with information relevant to the particular decision and how it affected them, rather than relying solely on a generic description of the system.

Right of Access

Under Section 26 of the Act, data subjects are entitled to access the personal data held about them, including data used as inputs to AI systems and the outputs, inferences, or scores generated about them.

Entities should be able to supply data subjects with information about what data is held and how it is used in AI processing, including, where processing is automated, the categories of data used and the logic applied in the decision or output.

Right to Rectification

Under Section 40 of the Act, data subjects have the right to have inaccurate personal data rectified. In the AI context, this includes the right to correct inaccurate input data and, where an AI output is based on inaccurate data, to have the output reviewed and corrected.

Entities should implement processes to promptly action rectification requests and to propagate corrections through AI systems where the inaccurate data has been used as an input.

Right to Erasure

Where the conditions for erasure under the Act are met, including where consent is withdrawn, where the data is no longer necessary for the purpose for which it was collected, or where processing is unlawful, entities should erase personal data from training datasets and implement appropriate measures to give effect to erasure requests.

Where erasure from trained model parameters is not technically feasible, entities should document the reasons and the basis for the feasibility assessment, assess and implement alternative mitigation measures, and maintain documented procedures for managing such requests.

Entities should address the technical challenges of erasure from AI models in their data protection governance frameworks.

Where an entity determines that erasure from trained model parameters is not technically feasible, it should nonetheless erase the personal data from training datasets, feature stores, embeddings, caches and logs; suppress the personal data from the model's inputs and outputs by technical means; record the assessment and the compensating measures adopted; and inform the data subject of the outcome and of the right to lodge a complaint with the Office. An assertion of technical infeasibility should be assessed on the evidence and does not operate as a general exemption from the right to erasure.

Right to Restriction

Data subjects have the right to request restriction of the processing of their personal data, including processing by AI systems. This applies where they contest the accuracy of the data, where the processing is unlawful, but they oppose erasure, where the entity no longer needs the data but the data subject requires it for a legal claim, or where they have objected to the processing and verification is pending.

Entities should give effect to restriction requests promptly. During the restriction they should limit processing to storage, unless the data subject consents or the processing is needed for legal claims or to protect the rights of another person. They should also inform the data subject before lifting the restriction.

Right to Object

Data subjects have the right to object to the processing of their personal data for purposes of profiling or direct marketing, including profiling by AI systems.

Entities should give effect to objections promptly and should not continue profiling a data subject who has objected, unless compelling legitimate grounds for the processing override the interests of the data subject.

Right Not to Be Subject to Automated Decision-Making

Section 35 of the Act provides that a data subject has the right not to be subject to a decision based solely on automated processing, including profiling, which significantly affects the data subject. This right is of central importance in the AI context.

Entities whose AI systems produce decisions that significantly affect data subjects, including decisions on credit, employment, insurance, housing, healthcare, education, or access to public services, should:

- ensure that such decisions are not made solely by automated means without meaningful human involvement;
- inform data subjects that a decision has been made through automated processing;
- provide data subjects with the right to request human review of any automated decision that significantly affects them;
- provide data subjects, upon request, with a meaningful explanation of the logic of the automated decision and its consequences; and
- implement mechanisms for data subjects to contest automated decisions and to have the contest considered by a human with authority to reverse or modify the decision.

In line with Regulation 22 of the Data Protection (General) Regulations, 2021, entities whose AI systems produce decisions based solely on automated processing should additionally:

- ensure the prevention of errors in the automated decision-making;
- use appropriate mathematical or statistical procedures for the profiling or automated processing;
- put in place appropriate technical and organisational measures to correct inaccuracies and minimise the risk of errors; and
- process personal data in a manner that eliminates discriminatory effects and bias.

For the purposes of Section 35 of the Act, entities should further:

- ensure that the person conducting a human review has the competence, authority, information and time to reach an independent conclusion and to reverse or modify the decision;
- acknowledge a request for human review promptly and determine it within the timeframes prescribed by the Act and the Regulations;
- ensure that the data subject is not placed at a disadvantage, and that no irreversible step is taken in reliance on the contested decision, while a request for human review or a challenge is pending, unless the entity is required by law to act;
- provide the data subject with the reasons for the outcome of the review in plain language, together with the route of escalation and the right to complain to the Office; and
- maintain records of each request for human review, the identity and role of the reviewer, the information considered, the outcome and the time taken.

Where an agentic artificial intelligence system is capable of taking an action that produces legal effects concerning a data subject or similarly significantly affects them, including entering into a transaction, transferring funds, suspending an account or disclosing personal data to a third party, the entity should require documented human approval before the action is executed and should implement and test effective controls to pause, interrupt and reverse the action.

Example

Sokoni Insurance uses an AI model to assess claims and automatically decline those it classifies as high-risk for fraud. A claimant whose claim is declined by the AI model has the right under Section 35 of the Act to request human review of the decision.

Sokoni should:

- inform the claimant that the decision was made by an automated system;
- explain the principal factors that influenced the AI's determination; and
- ensure a qualified human reviewer examines the claim afresh before a final decision is communicated.

Example 2

A public hospital uses an AI system to prioritize patients for specialized treatment. Patients assessed as low priority are automatically deferred without further review.

A patient is not informed that the decision was made using AI, is not given access to the data or reasoning behind the outcome, and is not provided with a way to request human review.

This violates the rights of data subjects. The hospital fails to:

- inform the patient of the use of AI (Right to be Informed);
- provide access to data and inferences used (Right of Access); and
- allow the patient to request human review of a significant decision (Right not to be subject to automated decision-making).

Right to Data Portability

Under Section 38 of the Act, data subjects have the right to receive personal data they have provided to a controller in a structured, commonly used, and machine-readable format, and to transmit that data to another controller.

In the AI context, entities that generate personalised AI outputs based on data provided by the data subject should consider whether those outputs fall within the scope of the portability right and ensure they have the technical capability to produce portable data exports.

Example

A university uses an AI-powered learning platform to track student performance and generate personalized learning profiles based on coursework, assessments, and engagement. A student requests a copy of their data to transfer to another institution.

The university declines to provide the data, stating that the performance analytics and profiles are internal.

This violates the right to data portability. The university should provide the student with their personal data, including submitted coursework, assessment results, and learning profiles—in a structured, commonly used, and machine-readable format to enable transfer to another institution.

OBLIGATIONS OF DATA CONTROLLERS AND DATA PROCESSORS IN AI SYSTEMS

Registration

All entities that develop, deploy, or operate AI systems that process personal data of persons located in Kenya should register with the Office as data controllers or data processors, subject to the thresholds and procedures set out in the Data Protection (Registration of Data Controllers and Data Processors) Regulations, 2021.

Entities procuring AI-as-a-service products from third-party vendors should assess whether the vendor is acting as a data processor and, if so, should ensure the vendor is registered as required by the Act and enters into a compliant data processing agreement.

For further guidance on registration, the Office has published a Guidance Note on Registration of Data Controllers and Data Processors which is accessible at www.odpc.go.ke.

Privacy by Design and Default

Privacy by design and default, recognised under Section 41 of the Act and the General Regulations requires that data protection considerations be embedded into AI systems from the earliest stages of design and development, not addressed as an afterthought or add-on.

In the AI context, privacy by design obligations include:

- data minimisation at the data collection and feature selection stage;
- designing AI models to operate pseudonymised or aggregated data wherever feasible;
- implementing differential privacy or other privacy-enhancing techniques in AI training pipelines where technically appropriate;
- building in explainability and human review mechanisms as core system features;
- designing AI systems to default to the most privacy-protective settings available; and
- conducting privacy impact assessments at the system design stage.

Data Protection Impact Assessment (DPIA)

Section 31 of the Act and the General Regulations require a DPIA before commencing any processing activity that is likely to result in high risk to the rights and freedoms of data subjects.

AI systems that involve large-scale profiling, systematic monitoring, automated decision-making with significant effects, use of sensitive personal data, or processing of children's data should typically trigger the DPIA requirement.

A DPIA for an AI system should address:

- a description of the AI system, its purpose, the personal data processed, and the envisaged outputs and decisions;
- an assessment of the necessity and proportionality of the processing in relation to its purpose;

- an assessment of the impact of the artificial intelligence system on the rights and freedoms protected by the Constitution, including equality and non-discrimination, human dignity, and access to services;
- a record of the data used to train, fine-tune and evaluate the system, its provenance and the lawful basis relied upon;
- the results of accuracy, robustness, security and bias testing, including performance disaggregated by the groups the system may affect differently;
- a model card or equivalent record describing the intended purpose, limitations and known risks of the system;
- the arrangements for human review, contestability and the exercise of data subject rights;
- an assessment of the risks to data subjects, including discriminatory outputs, re-identification, data security risks, and the risk of erroneous or unexplainable automated decisions; and
- the measures proposed to address those risks, including technical controls, governance mechanisms, and data subject rights mechanisms.

A DPIA should be reviewed and, where necessary, revised before any material change to the artificial intelligence system, including retraining on new data, a change of purpose or of the population served, a material change in performance, the introduction of a new data source or tool, a change of model or provider, and extension of the system to a new sector or category of data subject. A DPIA should in any event be reviewed at intervals proportionate to the risk and, for high-risk artificial intelligence applications, at least annually.

Where a DPIA indicates a high residual risk that the entity is unable to mitigate, the entity should consult the Office before commencing the processing. Entities deploying artificial intelligence systems that affect the public at large are encouraged to publish a summary of the DPIA, omitting information that is commercially confidential or whose disclosure would prejudice security.

The Office has published a Guidance Note on Data Protection Impact Assessments accessible at www.odpc.go.ke. Entities should consult that Guidance Note when conducting DPIAs for AI systems.

Data Protection Officer (DPO)

The Act, under Section 24, provides that a data controller or data processor may designate or appoint a Data Protection Officer. The designation or appointment of a DPO is strongly advisable for entities whose core activities involve large-scale systematic monitoring of data subjects or large-scale processing of sensitive personal data.

Many AI deployments, including large-scale AI-driven customer monitoring, health AI, and AI-based credit assessment platforms make the designation or appointment of a DPO particularly advisable. The DPO should be involved in all matters relating to the protection of personal data in AI systems, including DPIA processes, AI system procurement decisions, and data subject rights responses.

Sensitive Personal Data and AI

The processing of sensitive personal data, including health data, biometric data, racial or ethnic origin, and data revealing sexual orientation in AI systems requires explicit consent or another specific legal basis permitting such processing.

Entities should not use sensitive personal data as AI training data or as inference inputs without a clear legal basis and without implementing heightened technical and organisational safeguards.

AI systems that generate inferences about sensitive personal data categories, for example, systems that infer health status or ethnicity from non-sensitive inputs may themselves produce sensitive personal data.

Entities should assess whether AI outputs constitute sensitive personal data and apply the corresponding legal standards.

Entities should assess and address the risk of proxy discrimination, that is, the use of variables that are not themselves sensitive personal data but that correlate closely with a protected characteristic, including place of residence, name, language, device type, educational institution and social network. The absence of a sensitive data field from the inputs does not establish that the system does not discriminate.

Children and AI Systems

Section 33 of the Act provides enhanced protections for children's personal data.

Entities deploying AI systems that are likely to be accessed by or that process data about children should implement age-appropriate design measures, including age verification mechanisms, and should obtain verifiable parental or guardian consent for the processing of children's data.

AI systems should not be used to profile children for commercial purposes or to make automated decisions that significantly affect children without appropriate human oversight and the involvement of the child's parent or guardian.

In addition, entities deploying artificial intelligence systems that are likely to be accessed by children or that process children's personal data should:

- assess and act in the best interests of the child at the design stage and record that assessment;
- refrain from profiling children for commercial purposes, including behavioural advertising and engagement optimisation, and from design techniques that exploit a child's inexperience or susceptibility to influence;
- refrain from deploying emotion recognition or emotion inference in respect of children in education, childcare and comparable settings;
- refrain from subjecting a child to a decision based solely on automated processing that produces legal effects concerning the child or similarly significantly affects the child;
- apply the most privacy-protective settings by default and limit collection to what is strictly necessary;
- conduct a DPIA for any such artificial intelligence system; and
- provide a clear route by which a child, parent or guardian may raise a concern and obtain a response.

Entities should apply comparable safeguards where an artificial intelligence system processes the personal data of other persons who may be at heightened risk, including older persons, persons with disabilities, refugees and asylum seekers, and persons in situations of economic vulnerability. In

relation to adults with disabilities, entities should respect the legal capacity of the person and support the exercise of their rights rather than defaulting to substituted decision-making.

Notification and Communication of a Breach

Entities should notify the Office of any personal data breach affecting AI systems, including unauthorised access to training datasets, model extraction attacks, or adversarial manipulation of AI outputs within seventy-two (72) hours of becoming aware of the breach, where the breach is likely to result in risk to data subjects.

Where the breach is likely to result in high risk to data subjects, affected individuals should also be notified without undue delay.

Where a personal data breach occurs within a data processor's systems, including those of a third-party AI service provider, the data processor should notify the data controller within forty-eight (48) hours of becoming aware of the breach, and the data controller remains responsible for notification to the Office and, where required, communication to affected data subjects.

Fairness, Non-Discrimination and Accessibility

The Constitution and Section 25 of the Act require that personal data be processed fairly. An artificial intelligence system that produces materially different outcomes for persons based on a protected characteristic, without lawful justification, does not meet that requirement. Entities deploying artificial intelligence systems that produce outputs or decisions concerning individuals should:

- test the system for discriminatory outcomes before deployment and at intervals thereafter, using data representative of the population the system should affect;
- measure and record performance disaggregated by the characteristics relevant to the deployment, including sex, age, disability, ethnicity, language and geography, to the extent that this can be done lawfully and proportionately;
- document and act upon any material disparity identified, including by retraining the model, restricting the scope of the system or discontinuing its use;
- avoid the use of proxies for protected characteristics that are not necessary for the stated purpose; and
- retain records of testing and mitigation and make them available to the Office on request.

Entities should ensure that artificial intelligence systems, the notices describing them and the channels for exercising data subject rights are accessible. Where a system relies on a biometric, voice, facial or behavioural interface, the entity should provide an alternative means of access that does not disadvantage the data subject and should provide reasonable accommodation on request.

Example

A private school deploys an AI system to monitor student behaviour and performance using biometric data (facial recognition) and health-related information to identify at-risk students. The system processes data of minors without obtaining verifiable parental consent. Subsequently, a breach occurs at a third-party service provider, exposing student biometric and health data, but the school delays notifying affected parents and the regulator.

This raises multiple violations. The school fails to:

- establish a valid legal basis and implement heightened safeguards for processing sensitive personal data;
- obtain parental consent and apply appropriate protections for children's data; and
- notify the breach promptly and inform affected data subjects as required.

As a result, the school fails to meet its obligations in handling sensitive personal data, protecting children, and responding to data breaches.

Data Sharing and Third-Party AI Processors

Where an entity engages a third party to provide AI services involving the processing of personal data, including cloud-based AI platforms, AI-as-a-service providers, and AI model vendors, the third party may act as a data processor, a data controller, or a joint controller depending on the circumstances of the processing.

Entities should assess and document the role of each third party in the AI ecosystem. Where the third-party acts as a data processor, the entity should enter into a written data processing agreement with the AI processor that:

- restricts the processor to processing personal data only on the instructions of the controller;
- requires the processor to implement appropriate technical and organisational security measures;
- prohibits the processor from engaging sub-processors without prior authorisation;
- requires the processor to notify the controller without undue delay of any personal data breach, of any request received directly from a data subject, and of any request for access to personal data received from a public authority;
- states whether personal data may be used to train, fine-tune or otherwise improve the processor's own models, and prohibits such use in the absence of express written authorisation and a lawful basis;
- provides for audit and inspection rights, including the right to receive evidence of security testing, independent assurance and certification;
- provides for orderly termination, including the return or deletion of personal data, the export of records in a usable format, and continuity arrangements if the service is discontinued;
- requires the processor to assist the controller in fulfilling data subject rights requests and DPIA obligations; and
- requires the processor to delete or return all personal data upon termination of the agreement.

Before procuring an artificial intelligence system or service, entities should conduct and document due diligence addressing, at a minimum, the identity, location and corporate group of the provider and of any sub-processor; the location of processing and storage; the provenance and lawful basis of the data used to train the model; the security controls and independent assurance held by the

provider; the arrangements for assisting with data subject rights; the provider's incident notification commitments; and the arrangements for exit and portability.

Where the provider of an artificial intelligence system or model is established outside Kenya, entities should require the provider to designate a representative or contact point in Kenya through whom the Office and data subjects may be reached and should record the alternative arrangements made where this is not achievable.

Cross-Border Transfers of AI-Processed Data

Many AI systems involve cross-border data transfers, including the transfer of training data to offshore model developers, the use of cloud-based AI inference services, and the sharing of AI outputs with offshore recipients. Part VI of the Act restricts transfers of personal data outside Kenya to jurisdictions that provide adequate data protection, or where appropriate safeguards are in place.

Entities should assess and document the adequacy of data protection in any jurisdiction to which personal data is transferred in connection with AI processing and should implement contractual or other safeguards including binding corporate rules or contractual clauses where adequacy cannot be established. Entities may not transfer personal data to offshore AI processors without a lawful transfer basis.

Entities should carry out and document a transfer risk assessment before transferring personal data outside Kenya in connection with an artificial intelligence system. The assessment should consider the categories of personal data transferred; the purpose of the transfer, including whether the data should be used for model training; the law and practice of the destination jurisdiction, including the powers of public authorities to access the data; the safeguards relied upon, including contractual clauses and technical measures such as encryption, pseudonymisation or in-country processing; and the practical ability of a data subject to exercise rights and obtain redress.

The routing of an inference request to an artificial intelligence model hosted outside Kenya is a transfer of personal data and should be assessed accordingly, including where the transfer is transient. Entities should record the destination jurisdictions to which personal data is transferred in connection with each artificial intelligence system.

Example

A Kenyan bank deploys a cloud-based AI fraud detection system provided by an overseas vendor. Customer transaction data is shared with the vendor for processing, but no formal data processing agreement is in place. The data is transferred and processed on servers located outside Kenya without assessing the appropriate safeguards in the receiving jurisdiction or implementing appropriate safeguards.

This raises compliance failures. The bank fails to:

- enter into a data processing agreement governing the vendor's handling of personal data; and
- assess and establish a lawful basis for cross-border data transfer and implement appropriate safeguards.

As a result, the bank does not meet its obligations when engaging third-party AI processors and transferring personal data outside Kenya.

DATA PROTECTION AND THE AI LIFECYCLE

An AI system does not come into existence at a single moment: it passes through a series of distinct stages from initial conception to eventual decommissioning. Personal data may be involved at every one of these stages, and different data protection obligations attach to each stage. Entities should therefore approach compliance not as a one-time exercise at the point of deployment, but as a continuous obligation that tracks the entire lifecycle of the AI system.

This section maps the principal data protection obligations of the Act and the General Regulations to each stage of the AI lifecycle. It applies to both entities that develop AI systems internally and those that procure AI systems from third-party vendors, who should apply equivalent scrutiny to the procurement and integration stages.

AI Lifecycle Stage	Principal Data Protection Obligations
Stage 1: Problem Definition and AI System Design	DPIA scoping; privacy by design; lawful basis identification; DPO engagement
Stage 2: Data Collection and Preparation	Lawfulness, fairness, transparency; data minimisation; purpose limitation; consent or other lawful basis for training data; data quality controls
Stage 3: Model Training and Development	Purpose limitation; storage limitation on training data; security of training pipelines; bias assessment; pseudonymisation or anonymisation where feasible
Stage 4: Testing, Validation, and Pre-Deployment Review	Full DPIA completion; accuracy and fairness testing; explainability review; human review mechanism testing; data subject rights mechanism testing
Stage 5: Deployment and Integration	Privacy notice update; registration with the Office; data processor agreements; cross-border transfer assessment; breach notification procedures activation
Stage 6: Operation, Monitoring, and Maintenance	Ongoing accuracy and fairness monitoring; data subject rights fulfilment; breach detection and notification; periodic DPIA review; audit trail maintenance

Stage 7: Decommissioning and Data Disposal	Secure deletion of training datasets and operational data; erasure from model parameters where feasible; retention schedule compliance

Stage 1: Problem Definition and AI System Design

Data protection obligations begin before any personal data is processed. At the problem definition and system design stage, entities should:

- define the purpose of the AI system with sufficient specificity to enable lawful basis identification and DPIA scoping;
- engage the DPO at the earliest practicable stage, where one has been designated;
- identify whether the proposed AI system falls within a high-risk category requiring a mandatory DPIA;
- determine, at the architectural level, whether the AI system can be designed to operate without personal data, or with minimised, pseudonymised, or synthetic data; and
- embed privacy by design requirements in the AI system specification, including explainability, human review mechanisms, and data subject rights interfaces.

The privacy-by-design principle requires that data protection is not retrofitted after system design is complete but is a foundational design constraint from the outset.

Entities that procure AI systems from vendors should include privacy-by-design specifications in procurement requirements and should not accept AI systems that cannot demonstrate compliance with the Act.

Example

Fahari Health, a private hospital network, plans to deploy an AI clinical decision-support system to assist physicians with patient triage. At the design stage, the hospital's DPO is engaged to assess whether the system can operate on pseudonymised patient records rather than fully identified records, and to scope a DPIA.

The AI vendor is required to provide documentation of the model architecture, the categories of data required for inference, and the explainability mechanisms built into the system before the procurement contract is signed.

Stage 2: Data Collection and Preparation

The collection and preparation of data for AI training is a processing activity governed in full by the Act. Entities should:

- identify the lawful basis for collecting and processing personal data as training data, applying the same rigour required for any other processing activity;

- ensure that personal data used as training data was collected with a purpose that is compatible with the AI training purpose, or obtain fresh legal authority;
- apply data minimisation: collect the minimum categories and volume of personal data sufficient for the training purpose, and prefer anonymised, pseudonymised, or synthetic data where the training objective can be met without full personal data;
- implement data quality controls to ensure training data is accurate, representative, and free from systematic biases that would produce discriminatory model outputs;
- maintain records identifying the source of training datasets, the categories of personal data they contain, the lawful basis relied upon, and the data subjects' reasonable expectations at the time of original collection; and
- record, for each dataset, its source, the date of acquisition, the licence or contractual terms governing its use and the lawful basis relied upon, and retain that record for the life of the artificial intelligence system;
- where training data is sourced from third parties, conduct due diligence to ensure that the third party collected the data lawfully and that its transfer to the entity for AI training is permissible.

Entities should not use personal data scraped from the internet or aggregated from public sources as AI training data without assessing whether the collection and use of that data for training purposes is lawful under the Act. The fact that personal data is publicly accessible does not, of itself, provide a lawful basis for processing it as AI training data.

Example

Savannah Analytics proposes to train a natural language processing model on social media posts. The fact that a post is publicly accessible does not, by itself, authorise its use for AI training. Before collecting or using the posts, the firm ought to identify a lawful basis under section 30 of the Act, assess whether the training purpose is compatible with the purpose for which the data was published, and consider the risks to the people concerned. It should use synthetic text or effectively anonymised data where these can achieve the training purpose. If identifiable posts are necessary, the firm should justify that use, apply appropriate safeguards and determine whether the scale and nature of the processing require a DPIA.

Stage 3: Model Training and Development

During model training, personal data is processed by computational systems that learn statistical patterns from it. Entities should:

- implement appropriate technical and organisational security measures for training infrastructure, including access controls, encryption of training datasets at rest and in transit, and audit logging of access to training data;

- apply storage limitation: training datasets should not be retained beyond the period necessary for the training activity; upon completion of training, training data should be deleted or anonymised unless a separate lawful basis for continued retention exists;
- conduct bias testing during model development to detect and mitigate the risk of discriminatory outputs, particularly where training data reflects historical patterns of discrimination based on sensitive personal data categories;
- apply privacy-enhancing techniques, including differential privacy, federated learning, or secure multi-party computation where technically feasible and appropriate to the scale and sensitivity of the training data; and
- document the model development process, including the training data used, the architecture of the model, the training parameters, and the results of bias and accuracy testing.

Where AI model development is performed by a third-party developer on behalf of the data controller, the developer is acting as a data processor. A compliant data processing agreement should be in place before any personal data is shared with the developer for training purposes.

Example

Developing a hospital triage model: A hospital commissions a developer to train an AI model that helps prioritise patients for clinical review. The hospital documents the training purpose and instructs the developer on how the data may be used under a data processing agreement. The developer restricts and logs access to the training environment, encrypts the data and assesses whether training can be carried out without pooling identifiable records. During development, the hospital and developer test accuracy across relevant patient groups and address material differences in performance. They document the data and tests used, then delete or anonymise the training dataset when it is no longer needed, unless continued retention has a separate lawful basis.

Stage 4: Testing, Validation, and Pre-Deployment Review

Before an AI system is deployed in a live environment where it should process personal data of actual data subjects or produce decisions affecting them, entities should complete a pre-deployment review that includes:

- finalisation and sign-off of the DPIA, including documentation of the residual risks assessed as acceptable and the measures adopted to mitigate risks assessed as unacceptable;
- technical testing of the human review mechanism required under Section 35 of the Act to ensure data subjects can effectively request and receive human review of automated decisions;
- testing of data subject rights interfaces, including access request fulfilment, rectification workflows, and erasure mechanisms;
- accuracy and fairness validation using test datasets that are representative of the real-world population that should be subject to the AI system's outputs, with specific attention to differential performance across demographic groups;

- explainability review: verification that the AI system can generate intelligible explanations of its principal outputs and decisions in terms accessible to affected data subjects; and
- DPO sign-off, where a DPO has been designated, confirming that data protection obligations have been assessed and addressed.

Where testing requires the use of personal data, entities should ensure that a lawful basis exists for processing that data for testing purposes and should apply the same data minimisation and security obligations that apply to training data. The use of synthetic or anonymised test data should be preferred where it can provide meaningful test results.

Example

Pre-deployment review of an AI loan assessment system: Before using an AI system to assess live loan applications, a lender tests it with representative data to check its accuracy and whether error rates or outcomes differ materially across applicant groups. It verifies that staff can review and, where appropriate, change an automated decision, and that an applicant can request a review, correct inaccurate information and receive a clear explanation of the decision. The lender also tests its access and erasure processes, finalises the DPIA where required, records any remaining risks and obtains DPO input where a DPO has been designated. If identifiable applicant data is needed for testing, the lender documents the lawful basis and protects that data as it would in live use.

Stage 5: Deployment and Integration

At the point of deployment, entities should ensure that all compliance infrastructure is operational and that the AI system is integrated into the entity's broader data protection governance framework. Deployment obligations include:

- updating privacy notices to disclose the deployment of the AI system, the categories of personal data processed, the nature of any automated decision-making or profiling, and data subject rights;
- confirming registration with the Office, or updating an existing registration to reflect the new AI processing activity;
- executing data processing agreements with all third-party AI service providers, cloud infrastructure providers, and any other processors engaged in connection with the AI system;
- assessing and documenting the lawful basis for any cross-border transfers of personal data made in connection with the AI system, and implementing contractual safeguards where required;
- activating breach detection, logging, and notification procedures specific to the AI system; and
- conducting staff training on the AI system's data protection obligations, including how to handle data subject rights requests related to the AI system.

Example

Harambee Bank deploys an AI-driven fraud detection system that monitors customer transactions in real time and automatically flags and suspends accounts suspected of fraudulent activity. Prior to deployment:

- the bank updates its privacy notice to disclose the AI monitoring system;
- signs a data processing agreement with the AI vendor;
- conducts a cross-border transfer assessment (the AI vendor processes data on servers in the European Union); and
- trains customer-facing staff on how to handle customer enquiries about automated account suspensions, including how to facilitate human review requests under Section 35 of the Act.

Stage 6: Operation, Monitoring, and Maintenance

Once deployed, AI systems should be subject to ongoing monitoring and governance. Data protection compliance is not a point-in-time obligation: AI systems change over time through retraining, fine-tuning, and exposure to new data distributions, and the risk profile of an AI system may change materially without a new deployment event. Operational obligations include:

- ongoing monitoring of the AI system's outputs for accuracy, fairness, and consistency with the stated purpose, including detection of model drift and discriminatory output patterns;
- periodic review and, where necessary, revision of the DPIA, at intervals proportionate to the risk profile of the AI system and triggered by any material change in the system, the data processed, or the deployment context;
- maintenance of records of processing activities, including logs of AI system decisions and the data inputs used, to the extent necessary to demonstrate accountability and to respond to data subject rights requests and regulatory enquiries;
- fulfilment of data subject rights requests relating to the AI system within the statutory timeframes, including access requests, rectification requests, erasure requests, and requests for human review of automated decisions;
- prompt detection, containment, and notification of personal data breaches involving the AI system;
- periodic algorithmic auditing by a suitably qualified internal or independent auditor to assess model performance, bias, and compliance with data protection obligations;
- maintenance of a documented AI change management procedure covering model retraining, material changes to training datasets, deployment of the AI system in new contexts, and changes of third-party AI providers, with reassessment of privacy risks and review of the DPIA before material changes take effect; and
- maintenance of a complete and tamper-evident log of the actions taken by any agentic artificial intelligence system, including the tools invoked, the external systems accessed, the personal data disclosed and the human approvals given;

- maintenance of a documented AI incident management and remediation procedure, including a register of AI-related incidents such as discriminatory outcomes, systemic inaccuracies, and privacy harms, together with the remedial action taken and the notifications made.

Entities are encouraged to establish an AI governance committee or equivalent body with clear responsibility for monitoring AI system compliance on an ongoing basis. The DPO, where designated, should be included in that governance structure with appropriate access to AI system documentation and outputs.

Example

Monitoring an AI insurance claims system: An insurer uses AI to help assess claims. After deployment, it monitors errors, approval patterns and complaints, and finds that claims from one region are being flagged for further review more often than comparable claims elsewhere. It investigates whether changes in the data or model have caused the disparity, pauses the affected use if necessary, and arranges human review of disputed decisions. The insurer records the findings and corrective action, tests the updated model before use, and reviews its DPIA considering the change. Its DPO participates in the review; an AI governance committee may provide additional oversight if the insurer has established one.

Stage 7: Decommissioning and Data Disposal

When an AI system is retired or replaced, entities should ensure that the decommissioning process addresses all personal data processed in connection with the system. Decommissioning obligations include:

- secure deletion or anonymisation of all personal data held in training datasets, test datasets, and operational data stores associated with the AI system, in accordance with the entity's documented retention schedule;
- assessment of whether trained model weights or parameters encode personal data in a form that is recoverable through model inversion or membership inference techniques, and implementation of appropriate erasure measures where this risk is present;
- deletion or secure archiving of logs of AI system outputs that contain or relate to personal data, subject to any statutory minimum retention obligations;
- termination or amendment of data processing agreements with third-party AI processors upon cessation of the processing relationship, and confirmation of data deletion by those processors; and
- updating the records of processing activities to record the cessation of the AI processing activity.

Where an AI system is replaced by a successor system, entities should treat the successor system as a new AI deployment and apply the full lifecycle compliance process from Stage 1, including a fresh DPIA assessment that addresses any differences in architecture, data processing, or deployment context between the predecessor and successor systems.

Example

Replacing an AI fraud detection system: A bank replaces its AI transaction monitoring tool with a new provider's system. Before retiring the old tool, it inventories the personal data in its training and test datasets, operational records, model and logs. It deletes data whose retention period has ended, securely archives record it should keep, and obtains confirmation that the former provider has deleted or returned data held on the bank's behalf. The bank assesses whether the retired model could reveal information about individuals and takes appropriate steps to address that risk. It records the system's closure and assesses the new tool's data flows and risks before deployment, including whether a fresh DPIA is required.

HIGH-RISK AI APPLICATIONS: SPECIFIC OBLIGATIONS

Certain AI applications, by reason of the nature of the processing, the categories of data involved, or the significance of the decisions produced, pose heightened risks to data subjects and attract additional obligations.

The following categories are considered high-risk for the purposes of this Guidance Note and require, at a minimum:

- mandatory DPIA;
- DPO involvement, subject to Section 24 of the Act and applicable legal requirements;
- meaningful human review of automated decisions; and
- regular algorithmic audit.

An artificial intelligence application is high-risk for the purposes of this Guidance Note where one or more of the following criteria are met:

- the output or decision determines, or materially influences, a person's access to credit, insurance, employment, education, housing, healthcare, social protection or another essential service;
- the system is used by a public body, or by a private body performing a public function, in a manner that affects a person's rights, entitlements or obligations;
- the system processes sensitive personal data or produces inferences about sensitive personal data;
- the system processes the personal data of children or of other persons at heightened risk;
- the system involves biometric identification, biometric categorisation or emotion inference;
- the system involves systematic monitoring of a publicly accessible place, a workplace or an educational institution;
- the system operates at a scale, or with a degree of automation, such that an error would affect many persons before it could be detected and corrected; or
- the system takes actions autonomously that are difficult or impossible to reverse.

High-Risk AI Application	Primary Risks	Specific Obligations
AI Credit Scoring and Financial Decisioning	Discriminatory profiling; exclusion of vulnerable groups; opaque adverse decisions on loan eligibility, limits, and terms	DPIA required; Section 35 human review mechanism mandatory; explanation of principal scoring factors to be provided upon request; regular bias audit of model outputs

AI in Health Diagnostics and Clinical Decision Support	Erroneous clinical recommendations; processing of health data at scale; re-identification from de-identified clinical datasets	DPIA required; Health data processed under explicit consent or specific legal basis; DPO highly advised; compliance with Digital Health Act.
AI-Enabled Biometric Recognition (Facial, Fingerprint, Voice)	Mass surveillance; re-identification; processing of sensitive biometric data at scale	DPIA required; Explicit consent or statutory basis; DPO mandatory; strict access controls; prohibition on real-time biometric surveillance absent specific legal authority
AI in Employment Screening and Monitoring	Discriminatory filtering of candidates; covert monitoring; automated disciplinary decisions	DPIA required; Disclosure to candidates and employees of AI screening tools; human review of AI-recommended hiring or termination decisions; right to contest
AI in Education (Adaptive Learning, Assessment, Student Monitoring)	Profiling of children; automated grading or progression decisions; long-term data retention	Section 33 children's data protections apply; parental consent required; no commercial profiling of children; DPIA required
Generative AI Platforms (LLMs, Image Generation)	Training data privacy; reproduction of personal data in outputs; generation of misleading content about real persons	Privacy notice to disclose generative AI processing; training data audit for personal data compliance; output monitoring for personal data reproduction
AI in Law Enforcement and Public Safety	Discriminatory targeting; automated decisions affecting liberty; surveillance at scale	Specific statutory authority required; DPIA mandatory; independent oversight; restrictions on automated profiling for law enforcement purposes

In addition to the applications set out in the table above, the following are considered high-risk: artificial intelligence used to assess eligibility for public benefits and social protection; artificial intelligence used in tax administration, including risk scoring and compliance analytics; artificial

intelligence used in recruitment, performance management and workplace monitoring; artificial intelligence used in health triage, diagnosis and treatment prioritisation; artificial intelligence used in education, including admission, assessment, proctoring and behavioural monitoring; artificial intelligence used by law enforcement, including predictive policing, facial recognition and risk assessment; artificial intelligence used in the operation of digital identity systems and in linkage to national identity registers; and recommender systems operating at population scale.

An entity should reassess the risk classification of an artificial intelligence system where the system is retrained, its purpose or population changes, its performance materially changes, a new data source or tool is introduced, or an incident occurs, and in any event at least annually for systems classified as high-risk. Entities uncertain of the classification of a system may seek the guidance of the Office or apply for admission to a regulatory sandbox.

Example

AI Credit Scoring: Comparative Benchmark

A Kenyan digital lender deploys an AI credit scoring model using mobile transaction data and social metadata. Under the DPA, the lender should:

- disclose the AI scoring system in its privacy notice;
- identify a lawful basis for the profiling;
- conduct a DPIA;
- provide data subjects with the right to human review under Section 35; and
- be able to explain the principal factors in an adverse score on request.

PUBLIC SECTOR ARTIFICIAL INTELLIGENCE AND DIGITAL PUBLIC INFRASTRUCTURE

The Act applies to public and private entities alike. A public body deploying an artificial intelligence system is subject to the same obligations as a private entity and, because individuals ordinarily cannot decline to deal with the State, to a heightened standard of necessity, proportionality and transparency.

Public bodies, and private entities performing a public function or operating digital public infrastructure, should:

- identify the specific statutory or constitutional authority for the artificial intelligence processing and satisfy themselves that the processing is necessary and proportionate to a legitimate aim;
- conduct a DPIA before deployment and, for systems affecting the public at large, publish a summary of that assessment;
- publish a description of each artificial intelligence system used to make or support decisions about individuals, stating its purpose, the categories of personal data used, the lawful basis relied upon, the arrangements for human review and the route to challenge a decision, subject only to narrowly construed exceptions relating to national security and the prevention and detection of crime;
- publish or otherwise make available a summary of any arrangement under which personal data is shared with, or obtained from, another public body for use in an artificial intelligence system, stating the legal basis for the sharing;
- provide an accessible route, free of charge, by which an individual may obtain reasons for a decision, request human review and complain; and
- ensure independent oversight of the system, including internal audit and, where appropriate, external assurance.

Indiscriminate or untargeted biometric surveillance of the public is inconsistent with the Constitution and with the principles of the Act and should not be undertaken. Biometric identification in publicly accessible places may be undertaken only where it is authorised by law, targeted, time-limited, necessary and proportionate to a serious and specific purpose, subject to independent authorisation and oversight, and supported by a DPIA and by published information about its use.

Where an artificial intelligence system is used in connection with a national identity system or other digital public infrastructure, entities should ensure that the linkage of datasets is authorised by law and limited to the purpose for which it is authorised, that enrolment in or use of the system is not made a condition of access to an unrelated service without lawful basis, that a non-digital or non-biometric route to the service remains available, and that the system is subject to independent audit.

Artificial intelligence used to score, rank or profile individuals for the purpose of allocating public benefits or assessing tax compliance should be subject to a DPIA, to meaningful human review of any adverse determination, to published information about the criteria applied, and to an effective route of appeal.

Public bodies should not rely on a supplier's assertion of accuracy or fairness in the absence of independent verification and should incorporate the requirements of this Guidance Note into their procurement documentation.

PRE-PROCESSING CHECKLIST FOR AI SYSTEMS

Before commencing any processing of personal data using AI systems, organizations should complete the following checklist. Where any requirement is not met, the processing should not proceed until the deficiency is addressed, or appropriate safeguards are implemented and documented. Where high risks remain, the organization should conduct a Data Protection Impact Assessment and, where required, seek prior consultation with the Office of the Data Protection Commissioner before proceeding.

Compliance Area	Key Questions / Actions
Lawful Basis and Purpose	Have you clearly defined the purpose of the AI system? Have you identified and documented a valid lawful basis? Is the use of data for AI training compatible with the original purpose or supported by a new lawful basis?
Data Identification and Classification	What categories of personal data should be processed? Does the system involve sensitive personal data? Should it process data relating to children or vulnerable groups? Should inferred data or AI outputs constitute personal data?
Data Minimization and Design	Are you collecting only data that is necessary? Can anonymized, pseudonymized, or synthetic data be used? Has privacy by design been incorporated into the system?
Risk Assessment (DPIA)	Has a Data Protection Impact Assessment been conducted where required? Have risks such as bias, discrimination, re-identification, and security threats been identified? Are mitigation measures defined?
Transparency and Data Subject Rights	Have privacy notices been updated to disclose AI processing? Are mechanisms in place to support access, rectification, and human review? Can meaningful explanations of AI decisions be provided?
Governance and Accountability	Has responsibility for AI oversight been assigned? Is the Data Protection Officer involved where required? Are records of processing activities maintained?
Third-Party Processing	Should third-party AI providers process personal data? Are data processing agreements in place with appropriate safeguards?
Cross-Border Transfers	Should personal data be transferred outside Kenya? Has adequacy been assessed and appropriate safeguards implemented?

Security and System Readiness	Are appropriate security measures in place (access control, encryption, monitoring)? Has the system been tested for accuracy, fairness, and reliability? Are breach detection and notification procedures established?

AI GOVERNANCE FRAMEWORK

Entities that develop or deploy AI systems processing personal data should establish and document an AI governance framework, proportionate to the nature, scale, and risk of their AI processing, that specifies:

- the governance structures for AI oversight and reporting lines to senior management;
- the roles and responsibilities of each actor across the AI lifecycle, including the AI developer, AI provider, foundation model provider, AI deployer, and AI user, and the allocation of data controller, joint controller, and data processor responsibilities among them;
- the role of the DPO, where designated or appointed, in AI oversight, with appropriate access to AI system documentation and outputs;
- the criteria by which artificial intelligence systems are classified by risk and the process for reassessing that classification;
- the maintenance of an inventory or register of artificial intelligence systems, including systems obtained as a feature of a wider software product;
- the controls governing the use of third-party artificial intelligence services by staff and the prevention of unmanaged or shadow use;
- the arrangements for training staff, including reviewers of automated decisions, on their responsibilities under the Act;
- the retention of audit and assurance evidence for a period sufficient to demonstrate compliance and, in any event, for not less than the period during which the artificial intelligence system remains in use;
- the standards and frequency for algorithmic audits, including the qualification and independence of the auditor, the matters to be assessed (model performance, bias, drift, and compliance with data protection obligations), and the documentation of findings and remediation;
- the standards for meaningful human oversight of automated decisions, including the authority, competence, and information available to human reviewers and the escalation procedures for contested decisions; and
- the AI change management and incident management procedures described in this Guidance Note.

Algorithmic audits should be conducted before the deployment of a high-risk artificial intelligence system, at least annually thereafter, and following any material change or incident. The auditor should be independent of the team responsible for developing or operating the system. Audits should extend to artificial intelligence systems obtained from third parties to foundation models on which the entity's systems depend, and to recommender systems that rank or curate content or offers for data subjects. Where an entity is unable to audit a third-party model directly, it should obtain and assess the provider's own assurance documentation and record the limitations of that assurance.

GENERATIVE AI

Organisations that build, buy or use generative AI should comply with the Data Protection Act and Regulations whenever the system processes personal data. This includes tools that generate text,

images, audio, video or code. Privacy safeguards should be built into the system and enabled by default.

Depending on how the system is used, an organisation should:

- Identify and record a lawful basis for using personal data to train, fine-tune, test or supply information to the system. This includes data from public websites, third parties and user conversations.
- Limit the risk that the system reveals personal data from its training data, prompts, connected sources or memory. Test for this before launch and after retraining and review the risk regularly.
- Check for false or misleading statements about identifiable people, especially where someone may rely on the output to decide about them.
- Tell users when they are interacting with generative AI and make it clear when generated content could be mistaken for human-created content or affect someone's rights or interests.
- Provide a practical way for people to request access to, correction or deletion of their personal data, and to raise concerns about personal data appearing in an AI output.
- Assess risks such as memorisation, data leaks, re-identification, profiling and inaccurate outputs in a DPIA where one is required.
- Assign people to oversee the system, review concerning outputs, investigate incidents and ensure problems are corrected.
- Where a third-party provider is involved, set out each party's responsibilities in contracts and check that appropriate safeguards are in place.

If a system creates an artificial image, video or voice depicting a real, identifiable person, the organisation should make clear that the content was generated by AI and have a lawful basis for processing that person's data.

Organisations should also check whether a provider uses information entered by users to train or improve its models. They should configure the service to prevent that use unless there is a lawful basis and an express instruction permitting it and explain the arrangement to users.

AGENTIC ARTIFICIAL INTELLIGENCE AND RECOMMENDER SYSTEMS

Agentic artificial intelligence systems plan and execute sequences of actions, often by invoking external tools and services, with limited human intervention. Recommender systems determine the content, offers and options presented to data subjects. Both extend automated processing beyond the production of an output to the taking of action and the shaping of choice, and both require specific controls.

Entities that develop or deploy agentic artificial intelligence systems should:

(a) define and document the scope of the actions the system is permitted to take, the systems and data it may access and the limits of its authority, and configure the system so that it cannot exceed them;
(b) require documented human approval before the system takes a high-impact or irreversible action, including a financial transaction, the disclosure of personal data to a third party, the suspension or termination of a service, or a change to a person's entitlement;
(c) implement and test effective controls to pause, interrupt and roll back the actions of the system;
(d) maintain a complete and tamper-evident log of the actions taken, the tools invoked, the data accessed and the approvals given;
(e) assess and mitigate the risk that the system is manipulated through prompt injection, malicious content in retrieved sources or compromised tools; and
(f) identify the person accountable for the actions of the system; accountability for the processing remains with the data controller and cannot be transferred to the system or to its provider.

Entities that operate recommender systems should inform data subjects that content or offers are personalised, describe the principal categories of personal data and the criteria used to personalise them, provide an effective means to object to profiling and, where feasible, a non-personalised alternative, and assess the risk that personalisation produces discriminatory or exclusionary outcomes or exploits the vulnerability of a data subject. A recommender system that determines access to opportunities, including employment, credit, housing or education, should be treated as a high-risk artificial intelligence application.

ENFORCEMENT AND COMPLAINTS

The Office oversees the implementation and enforcement of the Act. It may receive and investigate complaints, conduct inspections and audits, issue enforcement and penalty notices, impose administrative fines, and make referrals for criminal prosecution where applicable. These powers extend to entities that process personal data through AI systems.

Entities using AI systems should have an accessible internal procedure for receiving, investigating and resolving data protection complaints. Data subjects should be able to raise a concern with the entity first and receive a timely response. A data subject who is dissatisfied with the outcome, or who does not receive a response, may lodge a complaint with the Office. The internal procedure does not limit the data subject's statutory right to complain to the Data Commissioner. Complaints may relate to, among other matters:

- failure to disclose AI processing in a privacy notice;
- use of AI in automated decision-making without providing a human review mechanism;
- failure to provide an explanation of an automated decision upon request;
- discriminatory AI profiling producing adverse outcomes;
- processing of sensitive personal data by an AI system without lawful basis;
- deployment of an artificial intelligence system by a public body without a lawful basis or without published information about its use;
- indiscriminate or untargeted biometric surveillance;
- failure to determine a request for human review within the prescribed time;
- use of personal data to train or fine-tune an artificial intelligence model without a lawful basis;
- failure to provide accessible rights channels or a non-biometric alternative means of access;
- failure to conduct a DPIA for a high-risk AI application; or
- unlawful cross-border transfer of personal data to an AI processor.

Entities found to have violated the Act in connection with AI processing are liable to enforcement action, including administrative fines and civil remedies, in accordance with the Data Protection (Complaints Handling Procedure and Enforcement) Regulations, 2021.

The Act applies equally to public and private entities. The enforcement powers of the Office are exercisable in respect of a public body in the same manner as in respect of a private entity.

The Office encourages entities to engage proactively with the Office on AI data protection questions, including by seeking advisory opinions and participating in ODPC-facilitated stakeholder engagement processes on AI governance. The Office may consider cooperation and proactive engagement among relevant factors during enforcement processes.

To lodge a complaint or seek further guidance, contact the Office at:

Office of the Data Protection Commissioner

Britam Towers, 12th Floor, Hospital Road, Upperhill, Nairobi

P.O. Box 30920-00100, Nairobi, Kenya

Email: complaint@odpc.go.ke (for lodging complaints) | info@odpc.go.ke / enquiries@odpc.go.ke (for enquiries and further guidance)

Telephone: 0207801800 | 0796954269 / 0752896867

Website: www.odpc.go.ke

ANNEX 1: COMPLIANCE CHECKLIST FOR AI SYSTEMS

The following checklist is a practical tool to assist entities in self-assessing compliance with data protection obligations applicable to AI systems. It does not constitute legal advice and does not substitute for a full DPIA or legal review.

Compliance Requirement	Action Required
Registration	Register as a data controller or processor with the ODPC at www.odpc.go.ke before deploying any AI system that processes personal data.
Privacy Notice	Update privacy notices to disclose: use of AI; categories of data processed; nature of automated decisions or profiling; significance and consequences for data subjects; and data subject rights, including the right to human review under Section 35.
Lawful Basis	Identify and document a lawful basis under Section 30 of the Act for each AI processing activity, including training, inference, and output generation.
Sensitive Personal Data	Identify whether the AI system processes sensitive personal data (including biometric, health, or inferred sensitive data). If so, ensure explicit consent or specific legal basis and implement heightened safeguards.
Children's Data	Assess whether the AI system may process children's data. If so, implement age verification, parental consent mechanisms, and age-appropriate design measures per Section 33 of the Act.
DPIA	Conduct a DPIA before deploying any high-risk AI application. Document the assessment and retain it for inspection by the Office. Refer to the ODPC DPIA Guidance Note.

DPO	Designate or appoint a DPO where the AI deployment involves large-scale systematic monitoring or large-scale processing of sensitive personal data.
Automated Decision-Making	Implement a human review mechanism for all AI systems that make or significantly contribute to decisions affecting data subjects. Ensure data subjects can request human review and receive an explanation of the decision logic.
Data Minimisation	Review AI training datasets and inference inputs to ensure only necessary data is collected and processed. Remove or anonymise data that is not required.
Data Processor Agreements	Ensure written data processing agreements are in place with all third-party AI vendors and cloud AI service providers, containing the mandatory provisions under the General Regulations.
Cross-Border Transfers	Assess whether AI processing involves cross-border data transfers. Ensure adequate legal basis for any transfer under Section 48 of the Act.
Breach Notification	Implement incident response procedures for AI-related data breaches. Ensure the Office is notified within 72 hours and affected data subjects are notified where required.
Ongoing Monitoring	Establish governance procedures for ongoing monitoring of AI systems for bias, drift, and emerging privacy risks. Conduct periodic algorithmic audits.
Staff Training	Train staff responsible for AI governance, including DPOs, data protection leads, and AI system administrators, on their obligations under this Guidance Note and the Act.

ANNEX 2: DPIA TRIGGER TABLE FOR AI USE CASES

The following table provides guidance on whether a DPIA is required for common AI use cases. A "Yes" indicates that a DPIA is mandatory before deployment. A "Review" indicates that a DPIA should be conducted unless a documented assessment determines the risk to be low. The use cases and obligations listed are illustrative and not exhaustive.

AI Use Case	DPIA Required?	Key Risk Factors
AI credit scoring for loan decisions	Yes	Large-scale profiling; automated decisions affecting financial rights; potential discriminatory outcomes
AI-driven health diagnostics using patient records	Yes	Sensitive health data at scale; clinical risk if AI errs; re-identification risk
Facial recognition for customer authentication	Yes	Biometric sensitive data; risk of mass surveillance; high re-identification risk
Generative AI chatbot with no personal data input	Review	Risk depends on whether personal data is incidentally processed in user inputs
AI email spam filter (enterprise internal)	Review	Lower risk if processing limited to internal communications without sensitive data
AI content recommendation engine (consumer platform)	Yes	Large-scale profiling of consumer preferences; potential sensitive data inference
AI student assessment and adaptive learning	Yes	Children's data; automated decisions affecting educational outcomes; long retention
AI in recruitment screening and candidate ranking	Yes	Profiling; potential discrimination; automated decisions affecting employment rights

AI fraud detection in payment systems	Yes	Large-scale transaction monitoring; false positives affecting access to financial services
AI predictive maintenance (industrial, no personal data)	No	No personal data processed
AI translation tool with anonymised inputs	Review	Low risk if inputs are genuinely anonymised; risk of re-identification from context
AI-based employee productivity monitoring	Yes	Systematic employee monitoring; potential sensitive inference; power imbalance

Entities uncertain whether a DPIA is required for a particular AI application are encouraged to contact the Office for guidance. The Office may, in appropriate cases, issue advisory opinions on DPIA requirements.

OFFICE OF THE DATA PROTECTION COMMISSIONER | www.odpc.go.ke
Britam Towers, 12th Floor, Hospital Road, Upperhill, Nairobi, Kenya
P.O. Box 30920-00100, G.P.O. Nairobi