

# ML Model Governance and Data Pipeline Policy (MLOps)

Document ID: IE-MGP-2025.01

Version: 1.0

Effective Date: 2025-11-01

Department: IT & Engineering - R&D, Data Science

## 1. Objective

To ensure that all Machine Learning models are developed, trained, deployed, and monitored in a secure, ethical, and reproducible manner.

## 2. Data Governance and Training Sets

- **Data Lineage:** All training data sets must be versioned, immutably stored, and traceable back to their source (e.g., anonymized client data, licensed public data).
- **Privacy:** Training data must adhere to the **Global Data Privacy Policy**, with mandatory anonymization or pseudonymization techniques applied before being used in the R&D pipeline.
- **Bias Detection:** Before training commences, data sets must be scanned for demographic or input feature bias using approved bias detection tools.

## 3. Model Lifecycle and Versioning

- **Experiment Tracking:** All experiments must be logged, tracking parameters, metrics (accuracy, fairness), and the specific data version used.
- **Model Registry:** Only approved models that have passed mandatory review (security, performance, bias testing) may be stored in the central model registry. Each model must have a unique, immutable version tag.
- **Deployment:** Model deployment must follow the standard CI/CD process (via containerization) and is managed via the MLOps pipeline, separate from the core application code deployment.

## 4. Monitoring and Drift Management

- **Inference Monitoring:** All deployed models must be monitored in real-time for:
  - **Performance Decay:** Drop in prediction accuracy or F1 score.
  - **Data Drift:** Changes in the input feature distribution compared to the training data.
  - **Bias Drift:** Changes in prediction fairness metrics across different demographic groups.
- **Retraining:** Automated alerting on drift or performance decay triggers an immediate **Human-in-the-Loop** review. If the threshold is breached, the model must be

quarantined and an automated retraining pipeline initiated with fresh data.