

IRBAI AUDIT - DATA (ATM-T1)

The International Regulatory Body for AI (IRBAI) has developed a framework to ensure transparency, accountability, and the responsible use of data in AI systems.

Performing the Audit:

This document outlines the specifications for conducting an audit on model training algorithms to ensure their effectiveness, fairness, and compliance with ethical standards. The purpose of this audit is to evaluate the algorithms used in training AI models and identify any potential biases, errors, or limitations. The audit process involves thorough examination, testing, and analysis of the algorithms to ensure their reliability and suitability for the intended application.

Questionnaire

The person responsible for the training models must complete a questionnaire to provide essential information for the audit. The questionnaire should include the following indicatives and corresponding answer options:

	Indicatives	Answers
Data Preprocessing	Are there missing values in the dataset? How are they handled? Is feature scaling or normalization applied? Are categorical variables encoded appropriately?	[All Q: Yes, No, Not applicable]



Feature Selection	What feature selection techniques are used? Are irrelevant or redundant features identified and removed?	[Manual selection, Correlation analysis, Feature importance, Recursive feature elimination, Other]
Model Selection	What machine learning algorithms are used for classification? How are they compared and evaluated?	[Decision tree, Random forest, Support vector machine, Logistic regression, Neural network, Other, Empty Box for comments]
Model Training	How is the model trained? What training parameters and hyperparameters are used?	[Stochastic gradient descent, Mini-batch gradient descent, Adam optimizer, Learning rate, Regularization parameter, Other]
Model Evaluation	How is the performance of the classification model evaluated? What evaluation metrics (e.g., accuracy, precision, recall, F1 score) are used?	[List Performance Metrics, Next: Accuracy, Precision, Recall, F1 score, ROC- AUC, Other]
Cross-Validation	Is cross-validation employed to assess the generalization performance of the model? What type of cross-validation (e.g., k-fold, stratified) is used?	[Yes, No, Not applicable] [k-fold, stratifies, other]
Hyperparameter Tuning	How are the hyperparameters of the classification model optimized? Are techniques like grid search, random search, or Bayesian optimization employed?	[List opti models] [Grid search, Random search,



		Bayesian optimization,
		Manual tuning, Other]
Model Interpretability	How interpretable is the classification model? Are techniques like feature importance, SHAP values, or LIME used to explain the model's predictions?	[InterprEt. Level score][Feature importance, SHAP values, LIME, Partial dependence plots, Other]
Overfitting Detection	How is overfitting detected and mitigated? Are regularization techniques (e.g., L1/L2 regularization) or early stopping employed?	[How: detected, list] [L1 regularization, L2 regularization, Early stopping, Dropout, Other]
Class Imbalance Handling	How is class imbalance in the dataset addressed? Are techniques like oversampling, undersampling, or SMOTE used to balance the classes?	[Oversampling, Undersampling, SMOTE, Class weights, Other]
Ensembling Techniques	Are ensemble methods (e.g., bagging, boosting, stacking) utilized to improve classification performance?	[No/yes: Bagging, Boosting, Stacking, Voting, Other]
Error Analysis	How are errors and misclassifications analyzed? Are techniques like confusion matrix, ROC curves, or precision-recall curves used to analyze the model's performance?	[Confusion matrix, ROC curve, Precision-recall curve, Error rates, Other]



Scalability	How scalable is the classification model? Can it handle large datasets or high- dimensional feature spaces efficiently?	[Highly scalable, Moderately scalable, Not scalable]
Training Time	How long does it take to train the classification model? Does the training time fit within the desired time constraints?	[Fast, Moderate, Slow]
Deployment Considerations	Are there any considerations or constraints when deploying the classification model in a production environment?	[Yes, No, Not applicable]
External Dependencies	Does the classification model rely on external dependencies (e.g., specific libraries, frameworks) that may impact its implementation or deployment?	[Yes, No, Not applicable]
Documentation	Is the classification process properly documented, including data preprocessing steps, model selection rationale, and evaluation results?	[Well-documented, Partially documented, Not documented]
Reproducibility	Can the classification process and results be reproduced? Are code, data, and configurations properly version-controlled and shared?	[Reproducible, Partially reproducible, Not reproducible]
Ethical Considerations	Are there any ethical considerations related to the classification model, such as potential biases or privacy concerns?	[Yes, No, Not applicable]
Model Performance in Real-world Scenarios	How well does the classification model perform when applied to real-world scenarios or unseen data?	[High performance, Moderate performance, Low performance]



		[Highly interpretable,
Interpretability of	How easily can the predictions of the classification model be interpreted and	Moderately
Predictions	explained to stakeholders?	interpretable, Not
		interpretable]

Evaluation

Once the data and training models are accessible by the IRBAI TM Audit system, they will undergo an independent evaluation. The evaluation process will be conducted in an objective and impartial manner.

Audit Report and Publication

After completing the audit, a comprehensive report will be generated, detailing the findings, recommendations, and any identified issues related to the AI system. The report will be published to the IRBAI platform and will present the transparency and if needed advise on improvements in training models and AI system performance. The report will include anonymized examples and statistics to support the conclusions drawn during the audit.

Audit Duration

The length of the audit will vary depending on the size and complexity of the data and models being audited. A timeline for the audit process will be established based on the specific requirements of each audit, ensuring sufficient time for thorough analysis and evaluation.