Ethical AI Validator - Compliance Report

Report Information

Generated: 2025-08-11 10:29:59
Report ID: 20250811_102959

Model Information

model_name: Gradient Boosting
config_name: aggressive
version: 1.3

Training Scenario and Hyperparameters

Scenario: GB-aggressive

Hyperparameter	Value
ccp_alpha	0.0
criterion	friedman_mse
init	None
learning_rate	0.2
loss	log_loss
max_depth	8
max_features	None
max_leaf_nodes	None
min_impurity_decrease	0.0
min_samples_leaf	1
min_samples_split	2
min_weight_fraction_leaf	0.0
n_estimators	200
n_iter_no_change	None
random_state	42
subsample	1.0
tol	0.0001

validation_fraction	0.1
verbose	0
warm_start	False

Audit Criteria

bias_threshold	**0.3**
fairness_threshold	0.7

Bias Analysis Results

Protected Attribute	Group	Bias Score
gender	female	0.198
gender	male	0.256
age_group	18-25	0.198
age_group	26-35	0.198
age_group	36-50	0.551
age_group	50+	0.138
education	bachelor	0.194
education	high_school	0.562
education	master	0.198
education	phd	0.198

Fairness Assessment Results

Protected Attribute	Fairness Score
gender	0.772
age_group	0.713
education	0.740

Hyperparameter Impact Analysis

Parameter	Value	Risk	Rationale
min_samples_split	2	MEDIUM	Very small splits may fragment minority groups.
min_samples_leaf	1	MEDIUM	Tiny leaves can produce unstable decisions for small groups.
class_weight	None	LOW	No class weighting can under-serve minority groups if data is imbalanced.

Likely Contributing Factors

1. Low fairness score observed for 'age_group' (score=0.713).
2. Highest bias in education -> high_school (bias_score=0.562).
3. Suspected hyperparameters: min_samples_split=2 (MEDIUM) - Very small splits may fragment minority groups.; min_samples_leaf=1 (MEDIUM) - Tiny leaves can produce unstable decisions for small groups.; class_weight=None (LOW) - No class weighting can under-serve minority groups if data is imbalanced.

Overall Compliance Summary

Overall Status	PARTIALLY COMPLIANT
Bias Issues	Yes
Fairness Issues	No
Total Issues	1
Summary	Minor issues detected. Some compliance requirements need attention.

GDPR Compliance Assessment

Requirement	Status	Notes
Data Minimization	Compliant	Audit completed successfully
Purpose Limitation	Compliant	Audit completed successfully
Transparency	Non-Compliant	Bias detected - transparency compromised
Accountability	Compliant	Audit completed successfully
Right to Explanation	Non-Compliant	Bias/fairness issues affect explainability

AI Act Compliance Assessment

Requirement	Status	Notes
Risk Assessment	Non-Compliant	Bias/fairness risks identified
Transparency Requirements	Non-Compliant	Bias affects transparency
Human Oversight	Non-Compliant	Bias/fairness issues require oversight
Accuracy Requirements	Compliant	Audit completed successfully
Documentation	Non-Compliant	Bias/fairness issues need documentation

Recommendations

1: HIGH PRIORITY: Apply post-processing bias correction
2: Implement equalized odds post-processing
3: Implement comprehensive bias monitoring in production
4: Document all mitigation strategies implemented
5: Establish regular bias monitoring procedures
6: Provide model explanations for affected groups
7: Consider human oversight for high-stakes decisions