DataGuardi
Enterprise Privacy Compliand

AI Model Risk Analysis Report

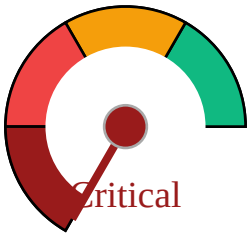
Generated on 2025-05-11 19:41:32

Executive Summary

This report presents the findings of an AI model risk assessment conducted on 2025-05-11 19:41:32. The assessment evaluated a ONNX model from Repository URL for privacy risks, bias concerns, and explainability issues. The analysis identified a total of 14 findings across multiple risk categories.

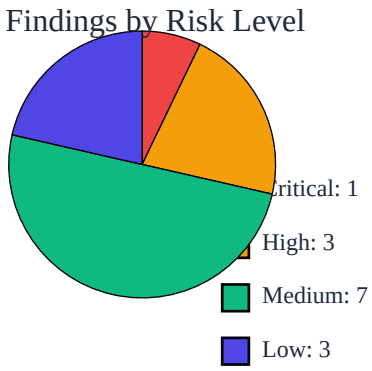
Scan ID:	AIMOD-20250511-249ee5
Model Type:	ONNX
Model Source:	Repository URL
Scan Date:	2025-05-11 19:41:32
Risk Score:	100/100
Total Findings:	14
Repository URL:	https://github.com/onnx/models
Branch:	main

Risk Assessment



Score: 100/100

Key Risk Metrics



Metric	Status	Risk Level
Personal Data in Model	✓ Detected	High
Bias/Fairness Issues	✓ Detected	High
Explainability Score	31/100	High

Detailed Findings

Open Source Compliance

ID	Type	Description	Risk Level
REPO-LICENSE-1b9d6b	License Detection	Repository has a Apache License 2.0 license	Low

Rights Management

ID	Type	Description	Risk Level
REPO-OPTOUT-6f1c29	Opt-Out Mechanism	Repository has a .gitignore file for excluding content	Low

Transparency

ID	Type	Description	Risk Level
REPO-DOCS-7d550	Documentation	Repository has documentation files that may contain attribution guidelines	Low

Architecture Analysis

ID	Type	Description	Risk Level
AIARCH-947746	Model Architecture	ONNX model architecture analyzed for privacy risks	Medium

Model Structure

ID	Type	Description	Risk Level
AIARCH-ONNX-04f556	Model Structure	ONNX model structure evaluated for exposed internal representations	Medium

PII Detection

ID	Type	Description	Risk Level

AIPII-TRAIN-56a7ae	Training Data PII	Model may contain unauthorized personal information in training data	High
AIPII-OUTPUT-e9bf3c	Output PII Leakage	Model may leak personal information in outputs through memorization	Critical

Model Bias

ID	Type	Description	Risk Level
AIBIAS-DI-827403	Disparate Impact	Model demonstrates potential disparate impact across protected groups	High

Explainability

ID	Type	Description	Risk Level
AIEXP-FI-749065	Feature Importance	Assessment of feature importance transparency	Medium
AIEXP-MI-a1788d	Model Interpretability	Overall model interpretability assessment for ONNX	Medium

GDPR Compliance

ID	Type	Description	Risk Level
AICOMP-c35aaa	Compliance Assessment	Model requires GDPR compliance assessment for Global	Medium

Technical Compliance

ID	Type	Description	Risk Level
AICOMP-ONNX-7b1171	ONNX Model Export Compliance	Assessment of ONNX model export for regulatory compliance	Medium

PII in Training

ID	Type	Description	Risk Level
AICOMP-TRAIN-fa0107	Training Data Assessment	Potential PII exposure in training data requires documentation	High

Transparency Requirements

ID	Type	Description	Risk Level
AICOMP-DOC-da2dda	Model Documentation	Assessment of model documentation for transparent use	Medium

Recommendations

1. Implement data minimization techniques to remove unnecessary personal data from the model
2. Conduct a Data Protection Impact Assessment (DPIA) for this AI model
3. Apply differential privacy to your training process
4. Implement bias mitigation techniques like reweighting or adversarial debiasing
5. Ensure diverse and representative training data
6. Use fairness constraints during model training
7. Enhance model transparency with feature importance visualization
8. Consider using more interpretable model architectures
9. Implement SHAP or LIME explanations for individual predictions
10. Prioritize addressing high and critical risk findings
11. Conduct regular AI model audits and ethical reviews

Conclusion

This AI model risk assessment identified 14 findings with a total risk score of 100/100. The model has personal data privacy concerns, exhibits bias issues, and has an explainability score of 31/100. By addressing the recommendations provided in this report, you can improve the model's compliance, fairness, and transparency.