Bias Audit Procedure & Reporting Template

A Practical Governance Template for Identifying, Measuring, Mitigating, and Documenting Algorithmic Bias

AI Governance Desk
www.aigovernancedesk.com

Document Overview

Purpose of This Template

This document provides a structured approach for conducting and documenting algorithmic bias audits across AI systems. It is designed to support teams in identifying disparities, evaluating fairness, applying mitigation strategies, and maintaining a clear record of audit outcomes within operational and governance contexts.

Regulatory and Governance Context

Bias auditing is increasingly expected as part of responsible AI governance and regulatory compliance, particularly for systems that influence high-impact decisions. This template aligns with emerging expectations under frameworks such as the EU AI Act and established risk management approaches, supporting transparent and defensible evaluation processes.

Scope of Application

This template is intended for use across a wide range of AI systems, including classification, ranking, and decision-support models. It is particularly relevant for use cases involving potential impacts on individuals or groups, such as lending, hiring, healthcare, insurance, and public-sector applications.

Intended Users

This document is designed for data scientists, machine learning engineers, AI governance and compliance teams, risk management professionals, product owners, and internal audit functions responsible for evaluating and managing algorithmic risk.

How to Use This Template

The template should be completed during the model development lifecycle and revisited during deployment and monitoring phases. Each section captures a specific component of the audit process, from data assessment to mitigation and documentation. Teams should adapt the structure to their internal processes while maintaining completeness and consistency across audits.

Audit Identification and Review Context

Bias Audit Identification

Audit Title _____

AI System / Model Name _____

Model Owner / Provider _____

Deployment Organization _____

Business Function / Domain _____

Use Case / Decision Context _____

Model Version _____

Audit Date _____

Review Period Covered _____

Audit Lead _____

This section anchors the audit to a clearly defined AI system, ownership structure, and evaluation timeframe. Accurate completion ensures traceability, accountability, and proper alignment with governance and regulatory expectations.

Audit Scope and Risk Classification

Audit Scope and Risk Context

System Type _____

Decision Type _____

High-Risk or Consequential Use Context _____

Primary Affected Populations _____

Protected Attributes Considered _____

Protected Attributes Unavailable / Legally Restricted _____

Risk Severity Level: **Low** □ **Medium** □ **High** □ **Critical** □

Potential Harm Domains _____

Scope Exclusions _____

Scope Rationale _____

This section defines the audit boundaries, identifies affected populations, and clarifies the level of risk associated with the AI system. Clear scoping is critical to ensure that bias evaluation is relevant, legally aligned, and focused on meaningful harm scenarios.

Data Assessment and Representativeness Review

Data Assessment

Training Data Description

Describe the dataset used for training, including size, structure, features, and target variables.

Validation / Test Data Description

Describe the datasets used for validation and testing, including any differences from training data.

Data Sources

List all data sources, including internal systems, third-party providers, or publicly available datasets.

Population Representation Review

Assess whether all relevant groups are adequately represented, including any known imbalances or gaps.

Missingness / Label Quality Review

Evaluate missing data patterns and the quality, reliability, and consistency of labels.

Sampling Concerns

Identify any sampling biases, selection effects, or collection issues that may affect fairness.

Proxy Variable Risks

Identify variables that may act as proxies for protected attributes (e.g., location, income, behavior patterns).

Data Limitations Summary

Summarize known limitations, constraints, and risks in the dataset.

This section evaluates the quality, representativeness, and limitations of the data used to develop and test the AI system. Weak data is one of the most common sources of bias, making this assessment critical for identifying fairness risks early in the audit process.

Protected Groups and Fairness Evaluation Setup

Fairness Evaluation Design

Group / Subgroup	Protected Attribute	Outcome & Metric	Notes / Justification

Analysis Basis

Describe how groups were defined and why they are relevant.

Harm Scenario

Describe the potential harm being evaluated across groups.

Metric Selection Rationale

Explain why the chosen fairness metrics are appropriate.

Data Availability Constraints

Document any limitations in measuring protected attributes or outcomes.

This section defines how fairness will be evaluated across relevant groups, including the selection of protected attributes, outcomes, and metrics. A well-defined evaluation design ensures that bias detection is meaningful, defensible, and aligned with real-world harm scenarios.

Fairness Metrics Results Table

Bias Measurement Results

Group / Subgroup	Metric Used	Baseline Result	Comparative Result	Threshold / Tolerance	Result Status ✓ | ⚠ | ✗	Reviewer Observations

Summary of Key Disparities Detected

This section captures the results of fairness evaluations across defined groups. It highlights disparities, compares them against acceptable thresholds, and provides a structured basis for identifying potential bias requiring mitigation.

Bias Diagnosis and Contributing Factors

Observed Disparity

Describe the disparity identified in the fairness results.

Likely Contributing Factors

Summarize the main factors suspected to influence the disparity.

Evidence Type

Specify how the issue was identified (e.g., statistical analysis, model inspection, stakeholder feedback).

Data-Related Causes

Identify issues related to data (e.g., imbalance, missing groups, label bias).

Model-Related Causes

Identify model behavior contributing to disparity (e.g., feature weighting, thresholds).

Evaluation-Related Causes

Identify issues in metrics or evaluation setup.

Deployment-Related Causes

Identify real-world or operational factors affecting outcomes.

Confidence in Diagnosis: Low □ Medium □ High □

Notes / Supporting Explanation

This section analyzes the root causes of observed disparities, linking outcomes to data, model, evaluation, and deployment factors. A structured diagnosis is essential for selecting effective and targeted mitigation strategies.

Mitigation Actions and Trade-Off Review

Mitigation Strategy and Trade-Off Analysis

Mitigation Option Technique applied (e.g., reweighting, thresholding)	Type of Mitigation Pre-processing / In-processing / Post-processing	Expected Fairness Impact Expected improvement	Expected Performance Impact Accuracy / latency / cost	Trade-Off Summary Key compromise	Decision Status Selected / Rejected / Under Review

Selected Mitigation Strategy

_____

Justification for Selection

_____

_____

Business / Regulatory Considerations

_____

_____

Decision Owner _____

Decision Date _____

Implementation Priority: **High** ☐ **Medium** ☐ **Low** ☐

This section documents the mitigation strategies considered to address identified bias, along with their expected impact and associated trade-offs. Clear documentation of decisions ensures transparency, accountability, and alignment with fairness, performance, and regulatory requirements.

Validation and Re-Testing

Post-Mitigation Validation

Metric Retested	Before Mitigation	After Mitigation	Result Interpretation	Residual Disparity	Status

Validation Date _____

Reviewer _____

Additional Testing Required: **Yes** □ **No** □

If yes, describe: _____

_____

Validation Summary Statement

_____

_____

_____

_____

This section evaluates whether mitigation strategies effectively reduced identified disparities. It compares pre- and post-mitigation results, highlights remaining risks, and ensures that fairness improvements are validated before deployment or continued use.

Residual Risk and Decision Record

Residual Risk Statement

Remaining Disparities

Describe any disparities that persist after mitigation.

Explanation for Residual Risk

Explain why these disparities remain (e.g., data limitations, trade-offs).

Severity Assessment: Low □ Medium □ High □ Critical □

Identify groups still affected.

Deployment Decision:

Approved □ **Approved with Conditions □** **Not Approved □**

Conditions for Continued Use

Restrictions / Controls Required

Escalation Required: Yes □ No □

If yes, specify:

Governance Decision Rationale

Provide justification for the final decision.

This section documents the residual risks that remain after mitigation and records the final governance decision regarding system deployment or continued use. It ensures accountability, transparency, and alignment with organizational and regulatory expectations.

Monitoring and Re-Audit Plan

Continuous Monitoring Plan

Monitoring Indicator	Description	Threshold / Trigger	Monitoring Frequency	Owner	Action if Triggered

Drift Detection Triggers

Define conditions that indicate model drift or fairness degradation.

Re-Audit Trigger Events

Specify when a full audit must be repeated (e.g., model update, data shift).

Responsible Monitoring Team

Reporting Channel

Incident Escalation Path

Monitoring Records Required

Specify logs, reports, or documentation to maintain.

This section establishes a continuous monitoring framework to ensure that fairness and performance are maintained over time. It defines indicators, thresholds, responsibilities, and re-audit conditions to support ongoing governance and risk management.

Bias Audit Reporting Summary

Executive Reporting Summary

Audit Objective
Summarize the purpose of the audit and what was evaluated.

Main Disparities Identified
Highlight the most significant fairness issues discovered.

Key Findings
Summarize the most important insights from the audit.

Mitigation Actions Taken
Describe actions implemented to address identified bias.

Residual Risks
Summarize remaining risks after mitigation.

Recommended Next Steps
Outline actions required moving forward.

Leadership Summary Statement
Provide a concise, high-level statement summarizing the audit outcome, risk posture, and decision readiness.

This section provides a high-level summary of the bias audit, highlighting key findings, mitigation actions, residual risks, and recommended next steps. It is designed for executive stakeholders to quickly understand the audit outcome and support informed decision-making.

Governance Review and Sign-Off

Review and Approval Record

Reviewer Role	Name	Department	Signature / Approval	Date

Final Audit Status:

Approved □ **Approved with Conditions** □ **Rework Required** □

Deployment Status:

Approved for Deployment □
Restricted Deployment □
Not Approved □

Next Review Date: _____

Comments / Final Remarks

_____

_____

_____

This section captures the formal review and approval of the bias audit, documenting accountability across responsible roles. It confirms the final audit outcome, deployment status, and any conditions or follow-up actions required.

Version Control and Document History

Version	Revision Date	Changes Made	Updated By	Reviewed By	Approval Status

Document Owner _____

Governance Department _____

Next Scheduled Review _____

This section tracks document revisions, updates, and approvals over time. It ensures traceability, accountability, and alignment with governance and compliance requirements.

About AI Governance Desk

Who We Are

AI Governance Desk provides structured, practical resources to support responsible AI development, evaluation, and deployment. Our tools are designed to help organizations identify risks, ensure fairness, and align with evolving regulatory and ethical standards in artificial intelligence systems.

Disclaimer

This document is provided for informational and operational support purposes only. It does not constitute legal, regulatory, or compliance advice. Organizations are responsible for ensuring that their AI systems meet applicable laws, standards, and internal governance requirements.

Usage and Intellectual Property

Contact Information

Website:
www.aigovernancedesk.com
Email:
contact@aigovernancedesk.com

This document is part of a structured AI governance framework designed to support transparency, accountability, and responsible AI system deployment.

Prepared By
AI Governance Desk

Document Version
v1.0

Publication Date: 18-03-2026