AI Bias Audit Framework

Document Purpose: This framework provides a structured approach to auditing AI systems used in health initiatives under the *Planetary Health Accord Implementation Framework*. It ensures AI technologies, such as those for diagnostics, resource allocation, or crisis response, are free from bias, culturally appropriate, and aligned with health equity principles. The framework is designed for AI developers, health policymakers, regional coordinators, community advocates, and Global Health Equity Council members to uphold transparency, fairness, and inclusivity.

Estimated Reading Time: 18 minutes

In this Guide:

- Overview

- Audit Scope and Objectives

- Audit Process

- Evaluation Criteria

- Stakeholder Involvement

- Cultural Competency and Equity Considerations

- Templates and Checklists

- Resources and Support

Audit Scope and Objectives

Scope

The framework applies to all AI systems deployed under the Accord, including:

- **Diagnostics**: AI tools for disease detection or risk assessment (e.g., predictive models for pandemics).

- **Resource Allocation**: Systems for distributing Global Health Solidarity Fund resources or medical supplies.
- **Crisis Response**: AI-driven analytics for emergency planning or mobile health unit deployment.
- **Health Equity Monitoring**: Dashboards tracking metrics like vaccination rates or maternal health.
- **Community Engagement**: AI chatbots or platforms for public health communication.

Objectives

- **Bias Detection**: Identify biases in data, algorithms, or outcomes that disproportionately harm marginalized groups (e.g., LMIC communities, indigenous populations).
- **Equity Assurance**: Ensure AI promotes universal health coverage and reduces disparities.
- **Cultural Alignment**: Verify AI respects indigenous knowledge and local cultural contexts.
- **Transparency**: Provide clear, accessible audit reports to stakeholders and the public.
- **Accountability**: Enforce corrective actions for non-compliant AI systems.

Checklist for Scope:

- ☐ Identify AI systems in use (e.g., diagnostics, resource allocation).
- ☐ Confirm systems fall under Accord governance.
- ☐ Document intended use cases and target populations.
- ☐ Assess potential equity impacts.

Audit Process

The audit process is a structured, multi-phase approach to ensure thorough evaluation and mitigation of biases.

Step 1: Pre-Audit Preparation (1-2 Weeks)

- **Objective**: Define audit scope and assemble team.
- **Actions**:

- Identify AI system and its use case (e.g., diagnostic tool for tuberculosis).

- Form audit team (5-7 members, see Stakeholder Involvement).

- Gather documentation: data sources, algorithm design, deployment protocols.

- Notify stakeholders (e.g., hub coordinators, community leaders).

- **Outcome**: Audit plan with timeline and responsibilities.

- **Timeline**: 1-2 weeks.

Step 2: Data and Algorithm Review (2-4 Weeks)

- **Objective**: Assess data quality and algorithm fairness.

- **Actions**:

  - Evaluate training data for representativeness (e.g., includes LMIC populations).

  - Check for biases in data collection (e.g., underreporting indigenous health outcomes).

  - Analyze algorithm logic for fairness (e.g., equal error rates across groups).

  - Use statistical tools (e.g., fairness metrics like demographic parity).

- **Outcome**: Report on data and algorithm biases.

- **Timeline**: 2-4 weeks.

Step 3: Outcome Analysis (2-3 Weeks)

- **Objective**: Examine AI outputs and real-world impacts.

- **Actions**:

  - Test AI outputs across diverse populations (e.g., urban vs. rural, indigenous vs. non-indigenous).

  - Collect feedback from end-users (e.g., Community Health Legions, patients).

  - Assess disparities in outcomes (e.g., misdiagnosis rates by gender or ethnicity).

  - Cross-reference with health equity dashboards.

- **Outcome**: Impact assessment report.

- **Timeline**: 2-3 weeks.

Step 4: Mitigation and Recommendations (1-2 Weeks)

- **Objective**: Address identified biases and propose solutions.

- **Actions**:
  - Develop mitigation strategies (e.g., retrain models, diversify data).
  - Consult indigenous and youth representatives for cultural alignment.
  - Recommend policy changes (e.g., ethical AI guidelines).
  - Submit recommendations to Regional Health Hubs or Global Health Equity Council.
- **Outcome**: Mitigation plan and policy recommendations.
- **Timeline**: 1-2 weeks.

Step 5: Reporting and Follow-Up (1-2 Weeks)

- **Objective**: Share findings and ensure accountability.
- **Actions**:
  - Publish audit report in accessible formats (e.g., infographics, multilingual).
  - Present findings at regional health forums or Global Biodiversity Forum.
  - Implement corrective actions within 30 days.
  - Schedule follow-up audit within 6-12 months.
- **Outcome**: Public report and action plan.
- **Timeline**: 1-2 weeks.

Template: Audit Workflow

```
**Audit ID**: [Unique Identifier]
**AI System**: [Name, Use Case]
**Start Date**: [DD-MM-YYYY]
**Audit Team**: [List Members]
**Step 1: Preparation**
- Scope: [System and Use Case]
- Timeline: [Dates]
**Step 2: Data/Algorithm Review**
- Findings: [Bias Issues]
**Step 3: Outcome Analysis**
- Disparities: [Group-Specific Impacts]
**Step 4: Mitigation**
- Actions: [Retraining, Policy Changes]
**Step 5: Reporting**
```

```
- Report Link: [URL]
- Follow-Up: [Date]
```

Evaluation Criteria

Audits use standardized criteria to assess AI systems for bias and equity.

- **Data Representativeness**:
    - Includes diverse populations (e.g., LMIC, indigenous, gender-diverse).
    - Accounts for regional health disparities.
    - Free from historical biases (e.g., skewed medical records).
- **Algorithm Fairness**:
    - Equal performance across groups (e.g., sensitivity, specificity).
    - No discriminatory weighting (e.g., prioritizing high-income regions).
    - Transparent model design (e.g., explainable AI).
- **Outcome Equity**:
    - No disproportionate harm to marginalized groups.
    - Aligns with health equity goals (e.g., universal coverage).
    - Validated by community feedback.
- **Cultural Appropriateness**:
    - Respects indigenous health practices and knowledge.
    - Adapts to local languages and cultural norms.
    - Avoids cultural stereotyping in outputs.
- **Transparency and Accountability**:
    - Clear documentation of data and model decisions.
    - Publicly accessible audit trails.
    - Enforceable corrective actions.

Checklist for Criteria:

- ☐ Verify data includes LMIC and indigenous populations.
- ☐ Test algorithm fairness across groups.
- ☐ Confirm outcomes align with equity goals.

- [ ] Ensure cultural respect in AI design.
- [ ] Document all processes transparently.

Stakeholder Involvement

Audits involve diverse stakeholders to ensure inclusivity and trust.

Audit Team Composition

- **Community Representative (1)**: From local health coalitions or patient advocacy groups.
- **Indigenous Advisor (1)**: Nominated by regional indigenous councils, with veto power on cultural issues.
- **Youth Representative (1)**: From Youth Advisory Boards, per *Youth Advisory Board Framework*.
- **Frontline Health Worker (1)**: Community Health Legion member or grassroots practitioner.
- **Technical Expert (1-2)**: AI ethics or public health specialist.
- **Regional Coordinator (1)**: From Regional Health Hubs, per *Regional Health Hub Implementation Guide*.

Engagement Methods

- **Community Feedback**:
  - Collect input via regional health forums and digital platforms.
  - Use participatory methods (e.g., focus groups, surveys).
- **Indigenous Consultation**:
  - Co-design audit criteria with indigenous advisors.
  - Respect cultural protocols (e.g., elder approvals).
- **Youth Participation**:
  - Involve Youth Advisory Boards in outcome analysis.
  - Host youth-led workshops to discuss AI impacts.
- **Transparency**:
  - Share draft findings with stakeholders for comment.

- Publish final reports in UN official languages and local dialects.

Template: Stakeholder Engagement Plan

```
**Audit ID**: [Unique Identifier]
**Stakeholder Group**: [Community/Indigenous/Youth/etc.]
**Role**: [Audit Team/Feedback Provider]
**Engagement Method**: [Forum/Survey/Workshop]
**Accessibility**: [Multilingual/Sign Language]
**Timeline**: [Dates]
**Outcome**: [Input Incorporated]
```

Cultural Competency and Equity Considerations

Audits prioritize cultural safety, indigenous sovereignty, and equity.

- **Indigenous Sovereignty**:
  - Grant veto power to indigenous advisors on cultural issues.
  - Protect traditional health knowledge from AI exploitation.
  - Use culturally appropriate evaluation methods.

- **Language Justice**:
  - Provide audit materials in regional and local languages.
  - Offer interpretation for meetings and reports.
  - Train auditors in multilingual communication.

- **Gender and Disability Inclusion**:
  - Ensure gender balance in audit teams.
  - Design accessible processes (e.g., braille, audio formats).

- **Marginalized Communities**:
  - Prioritize LMIC and marginalized group impacts in audits.
  - Address access barriers (e.g., low digital literacy).

- **Transparency**:
  - Use accessible reporting formats (e.g., infographics).
  - Allow public comment on audit outcomes.

Checklist for Equity:

- ☐ Include indigenous advisor with veto power.
- ☐ Translate materials into 3+ languages.
- ☐ Verify accessibility for disabilities.
- ☐ Assess impacts on LMIC communities.

Templates and Checklists

Template: Audit Report

```
**Audit ID**: [Unique Identifier]
**AI System**: [Name, Use Case]
**Date**: [DD-MM-YYYY]
**Audit Team**: [List Members]
**Findings**:
- Data: [Bias Issues]
- Algorithm: [Fairness Gaps]
- Outcomes: [Disparities]
**Cultural Issues**: [Indigenous or Local Concerns]
**Mitigation Plan**: [Actions, Timeline]
**Recommendations**: [Policy Changes]
**Follow-Up**: [Next Audit Date]
```

Template: Corrective Action Plan

```
**Audit ID**: [Unique Identifier]
**Issue**: [Bias or Disparity]
**Action**: [Retrain Model/Diversify Data]
**Responsible Party**: [Developer/Hub]
**Timeline**: [DD-MM-YYYY]
**Verification**: [Follow-Up Method]
**Stakeholder Input**: [Community/Youth Feedback]
```

Checklist: Audit Implementation

- ☐ Form diverse audit team.

- ☐ Document AI system data and design.

- ☐ Test outcomes across diverse groups.

- ☐ Develop and implement mitigation plan.

- ☐ Publish report in accessible formats.

- ☐ Schedule follow-up audit.

Resources and Support

- **Framework Documents**:
  - Governance Structure
  - Global Health Equity Council Setup Guide
  - Regional Health Hub Implementation Guide
  - Conflict Resolution Protocols
  - Youth Advisory Board Framework
- **Tools**:
  - Community Engagement Toolkit
  - Health Equity Dashboard Template
- **Support Channels**:
  - Email: [globalgovernanceframework@gmail.com]
  - Community Portal: [globalgovernanceframework.org/contact]
  - Quarterly AI Ethics Review Cycles for feedback.
- **Training Resources**:
  - Cultural Competency Training Module (Tools Library).
  - AI Ethics and Bias Audit Workshop (online, multilingual).

Call to Action: Start by identifying AI systems in use and forming an audit team with hub coordinators. Use the Community Engagement Toolkit to gather community input. Contact [globalgovernanceframework@gmail.com] for AI ethics training or to join audit pilot programs.

Cross-Reference Note: This framework integrates with the *Planetary Health Accord Implementation Framework*'s Governance Structure for oversight, Global Health Equity Council Setup Guide for Council roles, Regional Health Hub Implementation Guide for hub coordination, Conflict Resolution Protocols for disputes, and Youth Advisory Board Framework for youth involvement.