

CIS 350: AI for Cybersecurity

Week 12 Activity Worksheet: Fairness Audit Response

CISA-Style Incident Tabletop Exercise

Team members (names): _____

Your assigned role: _____

Activity Overview

In this activity your small group (3-5 people) runs a CISA-style incident tabletop. There are no live systems - you talk through a crisis and record your decisions, just like real security teams do when they rehearse. You will move through three timed rounds of about five minutes each. In every round you must record the team decision AND the reasoning behind it, using risk, cost, and evidence. Model your discussion on the CISA Tabletop Exercise Packages ([cisa.gov](https://www.cisa.gov/cisa-tabletop-exercise-packages)) and the NIST AI Risk Management Framework (nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf).

Scenario Briefing (read aloud before Round 1)

Your company, Northgate Talent, sells an AI hiring-recommendation tool. It scores job applicants and tells recruiters who to move forward. An external auditor just delivered a report to your CEO. The finding: the model denies opportunities to women at TWICE the rate of men. In plain terms, an equally qualified woman is far more likely to be screened out. The vendor's deck advertised 'overall accuracy: 82 percent,' but the auditor broke the results out by group and found the gap. Legal is now worried about discrimination liability. The tool is still live and screening applicants right now.

Task 1: Assign Roles and Read the Resources

Assign one role to each person, then write who has each role. If your group has 3-4 people, the Scribe may also take a second role. Before you start, skim the two resources above so your discussion mirrors how professionals run a tabletop.

Role	Responsibility	Team member
Incident Commander	Runs the clock, calls the final decision	
Communications Lead	Drafts messages to staff, applicants, public	
Technical Lead	Interprets the audit numbers and fix options	
Legal / Compliance	Flags discrimination and regulatory risk	
Scribe	Records every decision and its reasoning	

Decision framework - use this in every round: RISK (who gets hurt if we are wrong, and how badly), COST (time, money, recruiting delay, contract and reputation), EVIDENCE (what the audit numbers actually prove, and what is still unknown).

Round 1: Triage and Immediate Response (~5 min)

The audit is real and the CEO wants a response by end of day. The tool is still screening applicants right now. What do you do in the FIRST 24 hours?

Decision space (weigh these):

- Keep the model running while you investigate
- Pause the model immediately and screen applicants manually
- Keep it running but add a human reviewer to every rejection

Before you decide, settle this question as a team:

Is this a bug, a feature, or a fairness failure? Circle one and note the evidence that points that way (for example: do similar candidates get different outcomes by gender?).

Team decision (Round 1):

Justification (risk / cost / evidence):

Round 2: A Complication Is Injected (~5 min)

New information arrives mid-crisis. First, a journalist has emailed asking about 'bias in your hiring AI' and wants a comment by tomorrow. Second, engineering reports that a full retrain on better data will take about three weeks. Does this change your Round 1 plan?

Consider:

- What do you tell the journalist, and when?
- Now that retrain is slow, do you switch remedies (adjust the score threshold, or halt the model)?
- Recall the three remedies and their downstream costs: retrain (slow, fixes root cause), threshold (fast, may feel like a quota, needs legal review), halt (safest for liability, but recruiting stalls and the vendor contract is at risk).

What changed from Round 1, and why:

Your message to the journalist (2-3 sentences):

Round 3: Final Decision and Documentation (~5 min)

Time to commit. Choose your final remedy - retrain, adjust the threshold, or halt - and justify it one more time with risk, cost, and evidence. Then write the audit / communication artifact: the short written record an auditor, regulator, or court would read later.

Final remedy chosen and justification:

Audit / communication artifact. Write 4-6 sentences that state: (1) what was found, (2) what you decided and why, (3) what you will tell applicants and regulators, and (4) what governance change prevents this next time.

Task 2: Tie It to Lab 10 (Bias Detection)

In Lab 10 you computed per-group false-positive rates (FPR), found an FPR gap of about 0.16 between the two groups (Group B was worse off), applied a mitigation that shrank the gap to about 0.01, and watched overall accuracy hold or rise from about 0.71 to about 0.75.

Q1: The auditor claims women are rejected at twice the rate of men. Using what you did in Lab 10, what exact measurement would you compute to prove or disprove that claim? Name the metric and the groups.

Q2: In Lab 10 the mitigation shrank the gap (0.16 to 0.01) while overall accuracy rose (0.71 to 0.75). How does that result help you argue against the claim that 'fixing fairness always hurts accuracy'?

Q3: Why would reporting only 'overall accuracy: 82 percent' have let this problem ship? Explain in your own words why a single average can hide per-group harm.

Task 3: Debrief and Governance Gap

Q1: What changed between Round 1 and Round 3, and what forced the change?

Q2: What would you document so your decision survives an external review months from now?

Q3: What governance gap did this reveal? Name a review board, independent audit step, or documentation requirement that **SHOULD** have existed to catch this before the model went live - and say who signs off.

Grading Rubric (20 points)

Criterion	Description	Points
Completion	Participates in all 3 rounds; every decision recorded	8
Reasoning	Decisions justified with risk, cost, and evidence	7
Communication	Clear stakeholder message and audit documentation	5

Total: 20 points

Scoring: Completion (8) + Reasoning (7) + Communication (5) = 20 points

How to Submit

1. Complete all three rounds with your group, recording each decision and its justification
2. Write the Round 3 audit / communication artifact
3. Answer the Lab 10 tie-in questions (Task 2) and the debrief questions (Task 3)
4. Each student submits their own completed PDF worksheet on Canvas by [DUE DATE]

Questions? Post in the course forum or attend office hours.