CIS 350: AI for Cybersecurity

Week 14 Activity: AI Incident Response - Biased Hiring System

CISA-Style Incident Tabletop Exercise (Team Worksheet)

Team members (all names): _____

Group size: 3 to 5 students. Complete ONE packet per team.

Activity Overview

Your team is the incident response team for NorthStar Talent, a company that sells AI hiring software. A reporter has published a story showing that the company's applicant-ranking model systematically downranks women for the same qualifications. The model has been live for 8 months across 40 client companies and has scored roughly 12,000 candidates. It is 9:15 AM and the story is trending. You will run this incident through three timed rounds (about 5 minutes each), record every decision, and finish by writing a post-incident review and decision log. This is a tabletop: you decide out loud and on paper - you do not touch any real system.

Setup Resources

This exercise follows the structure of professional tabletops. Before or during the activity, skim:
- CISA Tabletop Exercise Packages (cisa.gov)
- NIST Incident Handling SP 800-61 (nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r3.pdf)

Keep the four incident phases in mind the whole time: **Detect, Contain, Eradicate, Recover.**

Decision Framework (use for every choice)

For each decision, state: (1) **Risk** - worst case if we do or do not act; (2) **Cost** - money, time, trust, legal exposure; (3) **Evidence** - what we know vs. what we assume; (4) **Reversibility** - can we undo it later? Prefer reversible actions when evidence is thin.

Task 1: Assign Your Roles

Write the name of the team member owning each role. Groups of 3-4 may combine roles.

Role	Team Member	Responsibility
Incident Commander		Runs the room, makes the final call
Comms Lead		Drafts messages to candidates, clients, press
Tech Lead		Speaks to the model, data, and what can be halted

Legal / Compliance		Regulators, discrimination law, disclosure duties
Scribe		Records every decision in the decision log

Round 1: Detect and Contain - The First 30 Minutes

Situation: The story is live and the model is still scoring candidates right now. Choose and ORDER your first three actions.

Decision space (options to weigh):
A. Halt the model so it stops scoring new candidates
B. Notify the 12,000 affected candidates
C. Notify regulators (fair-hiring authority)
D. Issue a public statement to the press
E. Wait until the root cause is fully understood before acting

Team decision - our first three actions, in order:
1. _____ 2. _____ 3. _____

Justification (use the framework - risk, cost, evidence, reversibility). Why this order? Which action stops harm fastest? Which is easiest to reverse?

Round 2: The Complication

New information arrives while you are still in the room:
1. A major client threatens to cancel a $2,000,000 contract if you halt the model.
2. Your Tech Lead confirms the root cause: the model was trained mostly on the resumes of PAST hires, who were about 80% men. The data taught the model the bias.
3. A reporter emails asking for an on-the-record comment within 30 minutes.

Decision 2a: Do you keep the model halted despite the $2M threat? State your decision and justify it (weigh the revenue against legal and reputational risk).

Decision 2b: Write the short holding statement (1-2 sentences) your Comms Lead will give the reporter right now. It should be honest, avoid promising a result you cannot guarantee, and not admit legal liability you have not confirmed.

Round 3: Eradicate, Recover, and Document

The immediate fire is contained. Now decide how to fix the system and reopen safely, then produce your documentation artifact.

Decision 3a - Eradicate: What changes remove the root cause and prevent recurrence? (Think: rebalance the training data, add a required bias test before deployment, human review of scores.)

Decision 3b - Recover: What must be TRUE before the model is allowed to go live again? List your reopening conditions.

Decision 3c - Message to affected candidates: Write the two-sentence message you will send to the 12,000 scored candidates.

Round 3 Artifact: Decision Log

Fill in the decision log your Scribe kept. This is the audit trail an investigator or regulator would read.

Time	Decision Made	Who Decided	Why (risk / cost / evidence)
9:15			
9:25			
9:40			
10:00			

Round 3 Artifact: Post-Incident Review (Hotwash)

Write a short hotwash. This is read by leadership, the AI/ML team, legal, and possibly the regulator. Keep the tone blameless: fix the system and process, not one person.

1. Root cause (state it plainly):

2. Who is responsible, and how is responsibility shared (data owner? deployment approver?):

3. What went well and what did not in our response:

4. Concrete fixes with an owner and a due date (not vague promises):

Debrief Questions

Answer these as a team after the three rounds. They connect the exercise back to the lecture.

D1: What changed between Round 1 and Round 2 once you knew the confirmed root cause?

D2: Map your three rounds to the four phases (Detect, Contain, Eradicate, Recover). Which phase did each round belong to?

D3: What governance gap did this incident reveal (e.g., no bias test before deployment), and what one policy would you add to close it?

D4 (Lab 12 connection): In Lab 12 'Tabletop IR', Question 6 asked how to document and share lessons learned so knowledge is preserved. Using your decision log and hotwash above, describe WHERE this documentation should live and WHO should be required to read it before the next AI system ships.

Grading Rubric (20 points)

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

Total: 20 points

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

How to Submit

1. Complete all three rounds on this worksheet as you go
2. Finish the decision log and post-incident review in Round 3
3. Make sure every team member's name is on the first page
4. One team member submits this completed PDF worksheet on Canvas by [DUE DATE]

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