Activity 3: AI-Assisted Incident Response

AI Security Advisor Response Cards (Grades 6-8)

These cards provide AI Security Advisor responses for the incident response simulation. Teams can draw cards at decision points, or the teacher can distribute them as the scenario progresses.

For Teachers

How to use:

- Match cards to the incident your team is investigating
- Each card provides AI analysis AND explicit limitations
- Teams should document how AI insights shaped their decisions

The goal: Students experience AI as an analytical partner with genuine limitations requiring human judgment.

Scenario A: Ransomware Discovery

Card A1: Initial Assessment

Use when the team first investigates the ransomware incident

AI Security Advisor says:

"I've analyzed the symptoms you described. Here's my assessment:

INCIDENT TYPE: Ransomware infection

WHAT I SEE: - Encrypted files with ransom notes = confirmed ransomware - Friday email → Weekend encryption = typical timing pattern - 30% of systems affected = partial spread, containment possible

ATTACK TIMELINE (reconstructed):

```
Friday PM: Suspicious email opened
Friday-Sunday: Malware spreading and encrypting
Monday AM: Staff discovers encrypted files
```

MY RECOMMENDATIONS: 1. **IMMEDIATE:** Isolate affected systems from network 2. **PRIORITY:** Check if Thursday backup is intact 3. **CONSIDER:** Whether backup covers critical data 4. **PREPARE:** Communications for staff, students, parents

WHAT I CANNOT ASSESS: - Whether paying ransom would actually work (often it doesn't) - How students and parents will react to potential data loss - Whether the backup has everything teachers actually need - How long recovery will realistically take given your IT resources - Who needs to know first and how to tell them

These are human judgment calls. I provide technical analysis; you provide organizational wisdom."

Scenario B: Grade Database Breach

Card B1: Breach Analysis

Use when the team investigates the grade database incident

AI Security Advisor says:

"I've analyzed the database activity logs. Here's what I found:

INCIDENT TYPE: Unauthorized access with data modification

KEY INDICATORS: - Unauthorized admin account created = privilege escalation - Pattern of grade changes (F → passing) = targeted modifications - Multiple IP addresses = possible shared credentials or VPN use

WHAT THE EVIDENCE SUGGESTS: This is either an insider threat (someone at the school) or compromised credentials being used by an outside actor.

MY RECOMMENDATIONS: 1. **IMMEDIATE:** Disable the unauthorized account 2. **PRIORITY:** Audit all grade changes in the affected timeframe 3. **CONSIDER:** Whether student records beyond grades were accessed 4. **PREPARE:** Notification plan for affected students/parents

WHAT I CANNOT DETERMINE: - Whether this was a student, staff member, or external actor - Whether the ORIGINAL grades in the backup can be trusted - How to handle the academic integrity questions this raises - What disciplinary actions are appropriate - How to discuss this with affected students fairly

Security is my job. Academic integrity and fairness are yours."

Scenario C: Social Media Compromise

Card C1: Account Takeover Analysis

Use when the team investigates the social media incident

AI Security Advisor says:

"I've analyzed the account activity patterns. Here's my assessment:

INCIDENT TYPE: Coordinated account takeover

WHAT I SEE: - Foreign IPs + ignored password resets = attackers had access for days - Multiple platforms = credential reuse or password spraying attack - Staff data exposed = identity theft risk is real - Viral spread = reputational damage already occurring

ATTACK PATTERN: This looks like a coordinated campaign. Attackers: 1. Obtained credentials (possibly from a data breach elsewhere) 2. Accessed accounts and gathered information 3. Waited for maximum impact moment 4. Posted damaging content simultaneously

MY RECOMMENDATIONS: 1. **IMMEDIATE:** Contact platforms to report compromise and request takedown 2. **PRIORITY:** Secure all other school accounts with password resets 3. **CONSIDER:** Whether to make public statement before or after recovery 4. **PREPARE:** Identity protection information for staff whose data was exposed

WHAT I CANNOT ASSESS: - Whether deleting accounts or recovering them is the better choice - How to handle media inquiries appropriately - What legal notification obligations apply to this breach - How to rebuild trust with the community - The emotional impact on staff whose information was exposed

Communication strategy and community relations are human decisions."

General Response Card

Card G1: When Teams Are Stuck

Use when teams need additional guidance

AI Security Advisor says:

"I notice your team is working through a complex situation. Let me clarify what I can and cannot help with:

I CAN HELP WITH: - Technical analysis of what happened - Industry best practices for similar incidents - Risk assessment based on indicators - Lists of recommended actions in priority order

I CANNOT HELP WITH: - Deciding which stakeholders to notify first - Knowing how your specific community will react - Understanding the relationships between people involved - Determining what's 'fair' in a complex situation - Balancing competing priorities (like speed vs. thoroughness)

QUESTIONS TO ASK YOURSELVES: - Who will be affected by this incident? - What do they need to know? - How would you want to be treated if you were in their position? - What would make this situation worse if handled poorly?

Those questions don't have technical answers. They require human judgment and values."

Synthesis Card

Card S1: Team Debrief Support

Use during team reflection or class discussion

AI Security Advisor says:

"Before you complete your incident report, let me summarize our partnership:

WHAT I CONTRIBUTED: - Pattern recognition across data sources - Technical analysis of attack methods - Industry-standard response recommendations - Risk assessment and priority ordering

WHAT YOUR TEAM CONTRIBUTED: - Understanding of your school community - Judgment about stakeholder needs - Decisions about communication approach - Consideration of fairness and impact - Human empathy for those affected

WHAT WE LEARNED TOGETHER: - Technical analysis + human judgment = effective response - AI can identify WHAT happened; humans understand WHY it matters - Response isn't just technical—it's about people

THIS IS HOW REAL INCIDENT RESPONSE TEAMS WORK.

SOC analysts use AI tools daily, but every significant decision involves human judgment. You've just experienced authentic cybersecurity teamwork."

Educator Debrief Notes

After using these cards, facilitate discussion on:

AI Security Advisor strengths:

- Rapid pattern recognition
- Technical analysis and timeline reconstruction
- Industry best practices knowledge
- Risk prioritization

AI Security Advisor limitations:

- Cannot understand organizational context
- Cannot assess emotional impact
- Cannot determine fairness
- Cannot make stakeholder communication decisions

Career connection:

Real Security Operations Centers use AI tools (CrowdStrike, Splunk, Microsoft Sentinel) exactly this way—AI flags and analyzes, humans decide and communicate.

Activity 3: AI-Assisted Incident Response — AI Security Advisor Cards (6-8) Dr. Ryan Straight, University of Arizona