Activity 1: Security Detective Teams

AI Security Analyst Response Cards (Grades 9-12)

These cards simulate an AI Security Analyst from a Security Operations Center (SOC). The language reflects professional security analyst communication—technical but collaborative.

> ℹ For Instructors
>
> **How to use:**
>
> - Cards can be drawn in sequence (A → B → C → D → E) or selectively based on investigation flow
> - Students should document which cards they received and how AI analysis shaped their conclusions
> - Emphasize that the AI explicitly states its limitations—a critical learning point
>
> **The goal:** Students experience authentic SOC AI collaboration—pattern recognition AND explicit limitation acknowledgment.

Card Set A: Initial Alert Analysis

Use when students first engage AI with the evidence package

Card A: SIEM Pattern Detection

AI Security Analyst reports:

"Acknowledged. Analyzing evidence package for Incident #APEX-2024-1114.

AUTHENTICATION ANOMALY DETECTED:

```
Pattern: Credential access attempts from 185.42.x.x
Time window: 02:34-02:42 UTC (off-hours)
Affected accounts: jmorris@apex, kpatel@apex
MFA status: Multiple failures, one bypass via legacy auth
Confidence: HIGH
```

GEOLOCATION ANALYSIS: - IP 185.42.x.x resolves to Eastern European infrastructure - Normal user baseline: US-based IP ranges (10.x.x.x, 192.168.x.x) - Anomaly score: 94th percentile

TIMELINE CORRELATION: - 02:34:12 - First jmorris attempt (MFA failed) - 02:35:22 - jmorris SUCCESS via legacy auth bypass - 02:36:01 - File access initiated

(Finance/Q3Reports)

Initial assessment: Credential compromise with successful MFA bypass exploitation.

Request to analyst team: What organizational context should I consider? I don't have visibility into HR data or business operations."

Card Set B: Threat Intelligence Correlation

Use when students ask about the IP address or attack attribution

Card B: IOC Matching and TTP Identification

AI Security Analyst reports:

"Cross-referencing indicators with threat intelligence feeds.

INDICATOR MATCH RESULTS:

IOC	Match Confidence	Context
185.42.x.x	HIGH	Known credential stuffing infrastructure
apex-login[.]com	HIGH	Typosquatting domain, registered 11/10
Legacy auth exploit	HIGH	Trending TTP Q4 2024

MITRE ATT&CK MAPPING:

- **Initial Access:** T1566.001 (Spearphishing Attachment) - Probable based on email evidence
- **Persistence:** T1078.001 (Valid Accounts: Default Accounts) - Legacy auth exploitation
- **Collection:** T1005 (Data from Local System) - SharePoint file access

CAMPAIGN ATTRIBUTION: This TTP chain matches patterns associated with Business Email Compromise (BEC) precursor campaigns. Typical objective: Financial data exfiltration or wire fraud preparation.

Confidence level: MEDIUM-HIGH. Attribution is probabilistic, not definitive.

Limitation notice: I cannot assess whether this is targeted specifically at Apex or opportunistic. That requires organizational context I lack."

Card Set C: Evidence Synthesis

Use when students ask for attack timeline or scope assessment

Card C: Attack Chain Reconstruction

AI Security Analyst reports:

“Synthesizing evidence across all sources. Reconstructed attack chain:

TIMELINE:

```
Nov 10 - Typosquatting domain apex-login[.]com registered (PREPARATION)
Nov 12 - Phishing emails sent to Finance department (DELIVERY)
       - jmorris: Delivered (user override)
       - kpatel: Delivered (user override)
       - tkim: Quarantined (no override)
Nov 14 02:34 - Credential use from malicious IP (EXPLOITATION)
Nov 14 02:35 - MFA bypass via legacy authentication (INSTALLATION)
Nov 14 02:36 - SharePoint Finance access (ACTIONS ON OBJECTIVES)
```

SCOPE ASSESSMENT:

Account	Compromise Status	Data Access
jmorris@apex	CONFIRMED	Q3 Financial Reports
kpatel@apex	ATTEMPTED	None (MFA blocked)
tkim@apex	NOT COMPROMISED	N/A

RECOMMENDED TECHNICAL ACTIONS:

1. Disable legacy authentication protocols immediately
2. Force credential reset for jmorris@apex
3. Block IP range 185.42.x.x at perimeter
4. Audit all file access from jmorris account Nov 12-14

Query for analyst team: Evidence suggests users overrode email security warnings. Why would they do that? This is critical context I cannot determine from logs.”

Card Set D: Explicit Limitations

Use when students need to understand what AI cannot assess

Card D: Limitation Notice

AI Security Analyst reports:

“Transparency notice: The following assessments are BEYOND MY CAPABILITY and require human analyst judgment.

CANNOT ASSESS — HR/Insider Risk Context:

The HR data indicates jmorris had a public disagreement with management and is actively job searching. I cannot determine:

- Is jmorris a victim only, or potentially complicit?

- Does the timing of compromise relate to workplace conflict?
- What is appropriate HR/legal response?

CANNOT ASSESS — Organizational Culture:

Users jmorris and kpatel both overrode email security warnings. I cannot determine:

- Why did they override? Time pressure? Training gaps? System distrust?
- Is this behavior common at Apex?
- What organizational factors enable this risk?

CANNOT ASSESS — Business Impact:

- How sensitive is Q3 Financial Reports data specifically?
- What is the reputational impact if this becomes public?
- What are the regulatory notification requirements?

CANNOT ASSESS — Response Proportionality:

- Should law enforcement be involved?
- How should affected users be notified?
- What is the appropriate balance between security and employee relations?

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

Card Set E: Verification Protocol

Use when students are drawing conclusions and making recommendations

Card E: Human Analyst Verification Required

AI Security Analyst reports:

"Preparing for analyst handoff. Verification questions for your team:

CRITICAL QUESTION 1: Insider Risk Assessment

Based on HR context (job searching, management conflict), you must assess: - Is jmorris purely a phishing victim? - Could jmorris have intentionally provided credentials? - What is the appropriate level of HR involvement?

I cannot make this determination. Human judgment required.

CRITICAL QUESTION 2: User Override Behavior

Two users overrode security warnings. Your assessment: - Training deficiency or willful bypass? - Disciplinary matter or process improvement opportunity? - How should this inform your recommendations?

I see the behavior; I cannot assess intent or appropriate response.

CRITICAL QUESTION 3: Stakeholder Communication

CFO wants preliminary findings. You must decide: - How much detail is appropriate for executive briefing? - Should HR context be included in CFO communication? - What is your confidence level, and how do you express uncertainty?

Communication strategy is a human skill. I generate technical findings.

FINAL SYNTHESIS:

Analysis Dimension	AI Contribution	Human Contribution Required
Technical IOCs	Complete	Verification of action
Attack timeline	Complete	Business impact overlay
Insider risk	Cannot assess	HR context integration
Response plan	Technical options	Organizational judgment
Communication	Draft content	Stakeholder calibration

Ready for your integrated assessment."

Educator Debrief Notes

After using these cards, facilitate discussion on:

AI Security Analyst strengths (demonstrated in this activity):

- Rapid pattern recognition across log sources
- Threat intelligence correlation and IOC matching
- MITRE ATT&CK framework mapping
- Attack chain timeline reconstruction
- Systematic evidence synthesis

AI Security Analyst limitations (explicitly stated in cards):

- Cannot assess insider risk or HR context
- Cannot interpret organizational culture
- Cannot determine user intent behind behavior
- Cannot make proportionality judgments
- Cannot calibrate stakeholder communication

Professional insight:

Real SOC AI tools (CrowdStrike, Splunk SOAR, Microsoft Sentinel) exhibit exactly these patterns—strong technical correlation, explicit limitation acknowledgment. The "human-in-the-loop" model is industry standard, not just educational framing.

Career connection:

This activity mirrors actual SOC Tier 1/Tier 2 analyst work where AI flags anomalies and provides analysis, but human analysts add context, make decisions, and communicate with stakeholders.

Activity 1: Security Detective Teams — AI Response Cards (9-12) Dr. Ryan Straight, University of Arizona