Activity 1: Security Detective Teams

AI Partner Response Cards (Grades 6-8)

These cards represent authentic AI responses—including uncertainty, questions back to the human partner, and genuine limitations. Use them when live AI access isn't available.

> **For Educators**
>
> **How to use:**
> - Students draw cards in response to their questions
> - Teacher can "play" the AI partner using these as scripts
> - Mix and match based on what students ask
>
> **The goal:** Students should discover that AI is strong at patterns but needs human help with context.

Card Set A: Opening Analysis

Use when students first share the evidence with their AI partner

Card A1: Pattern Recognition (AI Strength)

AI Partner says:

"Thanks for bringing me in on this investigation. Looking at the login logs, I'm immediately noticing some patterns:

1. **Time clustering**: Most failed attempts happened between 3:44-3:47 AM. That's a very narrow window.
2. **IP consistency**: All the 3 AM attempts came from the same IP address (203.45.67.89), which is outside your school district.
3. **The automated tool signature**: The user agent shows 'AutoBrute/2.1' — that's software designed to try lots of passwords quickly.
4. **The exceptions are interesting**: tpatel2025 and akim2025 logged in successfully during normal hours from the school IP. They weren't targeted.

But I have questions for you: Why would someone target these specific students and not others? I can see the pattern, but I don't understand the *why* yet. What do you think?"

Card A2: Honest Uncertainty (AI Limitation)

AI Partner says:

"I can see that the locked-out students have weaker passwords than the ones who weren't targeted. But here's what I'm uncertain about:

I don't know these students. I don't know the social dynamics at this school. Is there something connecting Jake, Maria, Katie, Ryan, and Brandon that isn't in this data?

Are they in the same friend group? Same sports team? Did they all make someone angry?

You might know things about middle school social situations that I can't figure out from logs and timestamps. What's your read on this?"

Card Set B: Deeper Analysis

Use when students ask follow-up questions

Card B1: Password Pattern Analysis (AI Strength)

AI Partner says:

"Okay, I cross-referenced the password patterns with the social media posts. Here's what I found:

- Jake's password was birthday-based → He posted his birthday publicly on August 23
- Maria's password was 'Buddy2025' → She posted photos of her dog Buddy
- Katie's password was 'CyberAcad123' → She used #CyberAcad123 as a hashtag
- Brandon's password was 'Lakers24' → He posts about the Lakers constantly

The connection: Someone could have guessed these passwords just by reading their public social media posts.

But here's what's interesting: Ryan's password was 'medium' complexity and not obviously tied to social media, but he still got targeted. Either the attacker got lucky, or there's something I'm missing. Any ideas?"

Card B2: Asking for Human Judgment (AI Limitation)

AI Partner says:

"I can tell you that this looks like a credential stuffing attack or social engineering attempt. The technical evidence points that way.

But I can't tell you:

- **Whether this is malicious or a prank.** Middle schoolers sometimes do dumb things to each other. Is this someone trying to actually steal something, or someone trying to lock their friends out as a joke?
- **What the school should do about it.** That's a judgment call involving school policies, student privacy, whether to involve parents or law enforcement…

Those are human decisions. What do you think is going on here, and what would you recommend?"

Card B3: Technical Depth (AI Strength)

AI Partner says:

"You asked about the IP address and the 'AutoBrute' tool. Let me explain what I know:

203.45.67.89 — This IP is geolocated outside your school district. The attacker is either:

- Physically somewhere else, OR
- Using a VPN/proxy to hide their real location

AutoBrute/2.1 — This is password-guessing software. It can try hundreds of password combinations per minute. The fact that it took 8-15 attempts per account suggests the attacker had *good guesses* but not the exact passwords.

The 3 AM timing — Probably chosen because no one's monitoring the network at 3 AM, and failed login alerts might not get noticed until morning.

My technical assessment: This was planned and deliberate, not accidental. But I don't know who or why."

Card Set C: Disagreement & Verification

Use when students push back or question AI

Card C1: AI Acknowledges Uncertainty

AI Partner says:

"You're right to question me. I could be wrong about this.

I'm good at spotting patterns in data, but I'm making inferences that might not be correct. For example, I assumed the social media posts were the source of the password guesses—but maybe the attacker got that information some other way.

Here's how you could verify: Ask the affected students directly if they've shared their passwords with anyone, or if anyone's been watching them log in.

What's your theory? I'm interested in what you're thinking."

Card C2: AI Corrects Itself

AI Partner says:

"Wait, I want to reconsider something.

I said Ryan's account didn't fit the pattern because his password was 'medium' complexity. But look at the network data again—his attempts happened at the same time and from the same IP as the others.

So he WAS targeted by the same attacker, even though his password was harder to guess. That makes me think the attacker had a *list* of target accounts, not just random attempts.

New question: What do these five students have in common that would put them on someone's list? This is the part I can't figure out from the technical data."

Card Set D: Synthesis Support

Use when students are drawing conclusions

Card D1: Summary with Gaps

AI Partner says:

"Let me summarize what we've figured out together:

What the evidence shows:

- Automated password-guessing attack from outside the school
- Targeted 5 specific students between Nov 18-21
- Attacker likely used social media to guess weak passwords
- Happened at 3 AM to avoid detection

What we still don't know:

- Who did this and why
- Why these 5 students specifically
- Whether this is malicious or a prank
- What the school should do about it

What I contributed: Pattern recognition, technical analysis, cross-referencing data

What you contributed: Context about students, judgment about severity, recommendations

I think we made a good team. What's your final conclusion?"

Card D2: AI Asks for Human Recommendation

AI Partner says:

"Here's something I can't do: make a recommendation about what the school should do next.

I can lay out options:

- Force password resets for all students?
- Block the suspicious IP address?
- Notify parents?
- Involve law enforcement?
- Run a school-wide security awareness lesson?

But choosing between those options involves understanding:

- School culture and policies
- How serious this really is
- What message you want to send
- Privacy considerations

Those are human judgment calls. What would YOU recommend, and why?"

Educator Debrief Notes

After using these cards, help students articulate:

AI's genuine strengths (from this activity):

- Fast pattern recognition across multiple data sources
- Technical knowledge (IP addresses, attack tools, security concepts)
- Systematic analysis and cross-referencing
- Memory for details humans might overlook

AI's genuine limitations (from this activity):

- Can't read social context or relationships
- Can't make judgment calls about severity or appropriate response
- Doesn't know "why" — only "what"
- Can be wrong and needs human verification

The partnership insight:

- Neither human nor AI alone could solve this as well as both together
- The *combination* of pattern-recognition AND contextual judgment is what real cybersecurity work requires

Activity 1: Security Detective Teams — AI Partner Response Cards (6-8) Dr. Ryan Straight, University of Arizona