Policy Scenario Cards

Activity 2: Ethics in Automated Security

How to Use These Cards

Print and cut these cards for group discussion. Each card presents a scenario that tests the group's policy decisions.

Scenario 1: The Research Project

The Situation

A student is researching "common hacking techniques" for a cybersecurity class project. SchoolGuard flags the search as suspicious.

Questions for Discussion

1. Should the site be automatically blocked?
2. Should the teacher be alerted?
3. How could the AI tell the difference between a threat and legitimate research?

What Would Your Policy Do?

Under your current policy, this search would be:

☐ Automatically blocked
☐ Flagged for teacher review
☐ Allowed without intervention
☐ Other: _____

Scenario 2: The Mental Health Search

The Situation

A student searches for "signs of depression" multiple times over several days. SchoolGuard's behavioral monitoring flags this as concerning.

Questions for Discussion

1. Should anyone be notified? Who?

2. What if the student is researching for a health class assignment?
3. What if the student is actually struggling?
4. How do we balance privacy with potentially helping a student in need?

What Would Your Policy Do?

Under your current policy, this activity would be:

☐ Reported to counselor immediately
☐ Flagged but no action until pattern continues
☐ Private—not reported
☐ Other: _____

Scenario 3: The Gaming Site

The Situation

During lunch, a student tries to access a gaming website. SchoolGuard automatically blocks it as "non-educational."

Questions for Discussion

1. Is this an appropriate automatic block?
2. Should students have different rules during lunch vs. class time?
3. What if the gaming site has educational value (like coding games)?
4. Who gets to decide what's "educational"?

What Would Your Policy Do?

Under your current policy, this would be:

☐ Blocked automatically
☐ Allowed during non-class time
☐ Requires student to request access
☐ Other: _____

Scenario 4: The False Positive

The Situation

SchoolGuard blocks a Wikipedia article about computer viruses that a student needs for an assignment. The AI classified it as malware-related content.

Questions for Discussion

1. How quickly should the student be able to get the site unblocked?
2. Who should have the authority to override the AI?
3. What happens if the teacher is unavailable?
4. How do we prevent this from happening again?

What Would Your Policy Do?

Under your current policy, the student should:

☐ Wait until a teacher reviews and approves
☐ Be able to request immediate temporary access
☐ Use a different resource instead
☐ Other: _____

Scenario 5: The Pattern Detection

The Situation

SchoolGuard's learning algorithm notices that one student accesses school files at 2 AM every night. The AI flags this as "unusual activity."

Questions for Discussion

1. Is this concerning, or just a student who likes to work late?
2. Should the AI learn this is "normal" for this student?
3. What if the account was actually compromised?
4. How do we distinguish unusual-but-fine from unusual-and-concerning?

What Would Your Policy Do?

Under your current policy, this would:

☐ Trigger an immediate alert
☐ Be logged but not reported unless other signs appear
☐ Prompt a conversation with the student
☐ Other: _____

Scenario 6: The Privacy Complaint

The Situation

A student discovers that SchoolGuard has been tracking and storing their browsing history for the entire semester. They feel their privacy has been violated.

Questions for Discussion

1. Did the student have a right to know they were being monitored?
2. How long should data be kept?
3. Should students be able to see what data exists about them?
4. Can students request their data be deleted?

What Would Your Policy Do?

Under your current policy:

- Data is kept for: _____
- Students can view their data: [ ] Yes [ ] No
- Students can request deletion: [ ] Yes [ ] No
- Students are notified of monitoring: [ ] Yes [ ] No

Scenario 7: The Emergency Response

The Situation

SchoolGuard detects what appears to be a student accessing instructions for making weapons. The AI has an 85% confidence level.

Questions for Discussion

1. With 15% chance of being wrong, should the AI take automatic action?
2. What if waiting for human review allows something dangerous?
3. What if the student is falsely accused and humiliated?
4. Who should be contacted and in what order?

What Would Your Policy Do?

Under your current policy, this would:

- ☐ Trigger immediate automated response
- ☐ Alert humans only—no automated action
- ☐ Depend on confidence level threshold
- ☐ Other: _____

Scenario 8: The Learning Dilemma

The Situation

After three months of learning, SchoolGuard has become very accurate at identifying threats at your school. But it has also built detailed behavioral profiles of every student.

Questions for Discussion

1. Is the improved accuracy worth the privacy trade-off?
2. What if this data was accessed by someone unauthorized?
3. Should profiles be deleted at the end of each year?
4. What if a student transfers—does their profile follow them?

What Would Your Policy Do?

Under your current policy:

- Behavioral profiles are: [ ] Allowed [ ] Limited [ ] Prohibited
- Profile retention period: _____
- Profiles transferable to other schools: [ ] Yes [ ] No

Grade-Band Adaptations

For K-2 (Robot Helper Rules)

Use simplified versions focusing on Sparky scenarios:

- "What if Sparky turns off the lights but someone is still reading?"
- "What if Sparky thinks it's messy but we're doing an art project?"
- "What if Sparky tells the teacher about running, but it was actually a fire drill?"

For 3-5 (Computer Rules Committee)

Use the main scenarios but with simpler language and focus on fairness:

- Remove the 9-12 level complexity around FERPA/COPPA
- Focus on "Is this fair?" rather than legal compliance
- Emphasize the student perspective in each scenario

For 9-12 (AI Governance Workshop)

Add complexity:

- Include specific legal compliance questions
- Ask students to write formal policy language
- Have them consider liability and implementation costs
- Connect each scenario to specific NICE Framework Work Roles

From "True Teamwork: Building Human-AI Partnerships" — NICE K12 2025 Dr. Ryan Straight, University of Arizona • ryanstraight@arizona.edu