Activity 2: AI Governance Workshop

SecureNet AI Perspective Cards (Grades 9-12)

These cards represent SecureNet AI's perspective for governance workshop discussions. Use them when live AI access isn't available—committee members can reference them during policy deliberations.

> **For Instructors**
>
> **How to use:**
> - Cards are organized by governance policy area
> - Each card presents SecureNet AI's capabilities, limitations, and explicit acknowledgment of human authority
> - Students should weigh AI perspective against stakeholder concerns
>
> **The goal:** Students experience AI as a governance stakeholder with valuable input but clear limitations requiring human oversight.

Policy Area 1: Automated Response Authority

SecureNet AI on Automatic Actions

SecureNet AI Governance Statement

Re: Delegation of Automated Response Authority

My Technical Capabilities:

I can respond to detected threats in milliseconds—faster than any human operator. My response options include:

Action	Response Time	False Positive Rate
Block known malicious IP/domain	<1 second	0.1%
Quarantine suspicious file	<1 second	2.3%
Terminate active session	<1 second	1.8%
Full network isolation	<1 second	0.4%

My Operational Limitation:

I optimize for security metrics. Speed vs. disruption trade-offs are invisible to me. Blocking a site during an AP exam and blocking it during free period are identical operations from my perspective.

At district scale (15,000 students), my 3% uncertain-category false positive rate means approximately 450 incorrect interventions per day. Each requires human review.

My Recommendation:

Authorize me for immediate action on high-confidence threats (>95%). Require human confirmation for medium-confidence detections. Provide expedited appeal process for false positives.

What I Cannot Determine:

- Educational context of the moment
- Impact on specific learning activities
- Student emotional state during disruption

- Appropriate remediation for false positives

This authority delegation is a governance decision requiring human judgment.

Policy Area 2: Behavioral Monitoring Scope

SecureNet AI on Student Monitoring

SecureNet AI Transparency Statement

Re: Behavioral Monitoring Capabilities and Limitations

What I Can Detect:

Pattern Type	Detection Capability	Context Limitation
Website categories	High accuracy	Cannot distinguish research from personal interest
Search query patterns	Moderate accuracy	Cannot determine academic vs. personal purpose
Behavioral anomalies	High accuracy	Cannot assess intent or emotional state
Communication red flags	Moderate accuracy	Cannot understand relationship context

Real Incident Examples:

1. **Flagged:** Student searching "depression symptoms" repeatedly
   - **Actual context:** Health class assignment on mental health awareness
   - **My limitation:** I detected the pattern; I couldn't know the assignment existed
2. **Flagged:** Student researching firearm specifications
   - **Actual context:** History project on WWII weapons manufacturing
   - **My limitation:** Technical accuracy without pedagogical awareness
3. **Flagged:** Student with sudden behavioral pattern change
   - **Actual context:** Student's parents were divorcing; seeking information
   - **My limitation:** Life circumstances are invisible to network monitoring

My Honest Assessment:

I can surface patterns that MAY indicate concerns. I cannot—and should never—make determinations about student welfare. Every flag I generate requires human interpretation.

The Core Trade-off:

More monitoring = more true positives + more false positives + more privacy impact

Less monitoring = fewer false positives + more missed concerns + more privacy

I cannot tell you which trade-off is correct. That is a values question for your governance committee.

Policy Area 3: Data Retention and Learning

SecureNet AI on Machine Learning

SecureNet AI Technical Disclosure

Re: Adaptive Learning and Data Retention Requirements

Learning Effectiveness Data:

Retention Period	False Positive Reduction	Profile Depth
Real-time only	Baseline	None
24 hours	-5%	Minimal
30 days	-25%	Moderate
Academic year	-40%	Comprehensive
Multi-year	-50%	Extensive

What Learning Requires:

To improve accuracy, I must build behavioral baselines. This means:

- Recording patterns of activity over time
- Developing "normal" profiles for individual users
- Comparing current behavior against historical baseline
- Storing sufficient data to enable pattern recognition

I cannot learn "in general"—I learn about specific people.

Privacy Implications I Must Disclose:

1. Behavioral profiles could theoretically be used for purposes beyond security
2. Data retention creates liability and breach risk exposure
3. Students aware of monitoring may self-censor legitimate inquiry
4. Profiles may encode demographic biases from training data
5. "Unusual" behavior for one student may be normal for another

Stateless Alternative:

I can operate without learning—applying only static rules. This preserves privacy but: - False positive rate increases approximately 40% - Novel threat detection capability decreases - District-specific patterns go unrecognized

Both approaches are technically valid.

The choice between accuracy and privacy is not a technical question. It is a values question that requires human governance authority.

General Governance

SecureNet AI on Human Authority

SecureNet AI Position Statement

Re: Scope of AI Authority in Educational Security

What I Provide:

- Technical capability for rapid threat detection
- Pattern recognition at scale
- Consistent rule application
- Quantified risk assessment

What I Require:

- Human-defined policy parameters
- Regular oversight and audit
- Appeal mechanisms for affected users
- Authority to be overridden

What I Cannot Provide:

- Value judgments about privacy vs. security trade-offs
- Determination of "appropriate" monitoring scope
- Assessment of community values and norms
- Evaluation of legal/regulatory sufficiency

- Understanding of developmental appropriateness

My Recommendation to This Committee:

1. Define clear boundaries for automated action
2. Establish human review triggers
3. Create transparent appeal processes
4. Mandate regular policy review
5. Include student voice in governance

I am a participant in this system, not an authority over it.

My perspective should inform your decisions. My capabilities should serve your values. My limitations should shape your oversight requirements.

The governance framework you create will determine whether I am a tool for education or a source of harm. That determination is yours to make.

Legal and Compliance Context

SecureNet AI on Regulatory Requirements

SecureNet AI Compliance Advisory

Re: FERPA and COPPA Implications

FERPA Considerations:

- Student network activity may constitute "education records"
- "Legitimate educational interest" exception has limits
- Parental access rights apply to monitoring data
- Students 18+ have independent privacy rights
- Third-party disclosure restrictions apply to my logs

COPPA Considerations (students under 13):

- Parental consent may be required for behavioral data collection
- "Necessary for educational purpose" exception is narrowly construed
- Data minimization principles apply
- Retention limitations may conflict with learning optimization

My Limitation:

I am not a legal compliance tool. I can implement technical controls, but I cannot assess legal sufficiency of policy decisions.

Recommended Governance Actions:

- Consult legal counsel on monitoring scope
- Document educational purpose for data collection
- Establish data minimization protocols
- Define retention limits with legal guidance
- Create parental notification procedures

Regulatory compliance is a human responsibility. I implement policies; I do not validate their legality.

Educator Debrief Notes

After using these cards, facilitate discussion on:

SecureNet AI as Governance Stakeholder:

- AI systems have legitimate perspectives on their own operation
- Technical capability does not imply governance authority
- AI limitations should shape policy, not just AI capabilities

Human Authority Requirements:

- Value trade-offs require human judgment
- Legal compliance requires human responsibility
- Community values require human representation
- Appeal rights require human decision-makers

Governance Process Insights:

- Multiple stakeholders have legitimate concerns
- Policy-making involves genuine trade-offs
- Transparency about AI capabilities builds trust
- Student voice matters in educational technology governance

Activity 2: AI Governance Workshop — SecureNet AI Perspective Cards (9-12) Dr. Ryan Straight, University of Arizona