**CYBR2100 Week 6 – AI Use in Security Operations**

# Policy Snippet (Final)

This AI use policy governs automated tools deployed in Hocking College’s Security Operations Center (SOC), including phishing detection, intrusion alerts, and anomaly monitoring. The purpose is to enhance awareness, speed, and accuracy while maintaining ethical and human-centered oversight. Users will see clear notices when an alert or classification is AI-generated, and each alert includes a short explanation of why it was triggered. Human analysts in the SOC are always in the loop, they can review, confirm, or override AI alerts through a standard verification console. Any student, staff, or administrator may appeal a flagged event through the IT Help Desk portal, with a response guaranteed within 24 hours. Data used by AI systems follows a minimum-necessary principle: logs are sanitized, retained for 30 days, and redacted of personal identifiers before review. SOC leadership tracks false positive and false negative rates (target ≤ 5%) and conducts quarterly retraining and fairness audits to ensure accuracy, transparency, and user trust.

# Controls & Metrics

1. **Human-in-the-loop Review**: All AI-generated alerts are verified by an analyst before escalation.  
2. **Subgroup Evaluation**: Quarterly bias and fairness testing across different user and event categories.  
3. **Appeals Proces**s: Users may dispute AI decisions through the IT Help Desk; response within 24 hours.  
4. **Data Retention Control**: Logs retained for 30 days, then securely deleted or redacted.  
5. **Performance Metrics**: Maintain false positives and false negatives below 5%; conduct quarterly retraining and audit reviews.

# Justification

These controls align with the ethical principles outlined in Chapter 11, which emphasize accountability, transparency, and fairness in AI-assisted decision systems. The human-in-the-loop and subgroup evaluations directly mitigate bias and over-reliance risks identified in the text. The metrics ensure measurable performance and fairness targets, while regular retraining maintains adaptive accuracy. Together, these practices operationalize responsible AI governance in security operations and protect user trust while minimizing unintended harm.

# Reflection

One trade-off I would revisit is the 30-day data retention period. While it aligns with privacy best practices, it may limit the ability to perform long-term trend analysis or retrospective investigations. A better balance might be a tiered retention system where anonymized data is retained longer for statistical review. I would also consider implementing monthly mini-audits in addition to quarterly reviews to ensure more continuous feedback on fairness and performance.