Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Human-AI Collaboration in Identity Security: When Should AI Decide?

Sujatha Lakshmi Narra
Independent Researcher, Atlanta, GA USA
Corresponding Author: Sujatha Lakshmi Narra, **E-mail**: sujathanarra87@gmail.com

| ABSTRACT

The integration of artificial intelligence into identity and access management presents both transformative opportunities and significant challenges for contemporary security frameworks. This article examines the critical question of decision authority allocation in AI-augmented security environments: determining when automated systems should independently make access determinations versus when human expertise remains essential. Through analysis of implementation case studies across financial services and healthcare sectors, the research identifies patterns of successful collaboration between algorithmic and human components of security ecosystems. The investigation reveals that optimal security outcomes emerge from thoughtfully designed frameworks that dynamically assign decision authority based on contextual risk factors, rather than static delegation models. Ethical dimensions receive particular attention, with privacy considerations, algorithmic fairness, and accountability mechanisms identified as critical success factors beyond technical implementation details. The article concludes with evidence-based recommendations for organizations implementing collaborative security models.

| KEYWORDS

Decision authority frameworks, Human-AI collaboration, Identity security, Algorithmic fairness, Contextual risk assessment

| ARTICLE INFORMATION

ACCEPTED: 12 June 2025 **PUBLISHED:** 02 July 2025 **DOI:** 10.32996/jcsts.2025.7.7.17

1. Introduction and Background

The rapid advancement of artificial intelligence (AI) technologies has revolutionized numerous sectors, with identity and access management (IAM) emerging as a particularly promising application domain. Recent systematic reviews of AI adoption in cybersecurity have documented significant growth, with implementation of AI-powered IAM solutions increasing annually since 2020. Organizations adopting these technologies have reported substantial reductions in security incidents compared to traditional rule-based approaches [1]. These systems can continuously monitor user behaviors, establish baseline patterns, and identify anomalies that may indicate security threats with unprecedented speed and accuracy. Contemporary security infrastructure implemented with machine learning components has demonstrated the capability to analyze thousands of authentication events per minute with high accuracy rates for known attack vectors [1].

However, this technological evolution raises fundamental questions about decision authority: When should AI systems autonomously make access decisions, and when should human expertise remain in the loop? Security frameworks implementing hybrid decision models that combine algorithmic assessment with human oversight report fewer false positives than fully automated systems, while simultaneously reducing incident response times compared to purely human-driven approaches [2].

2. Capabilities and Limitations of AI in Identity Security

AI-powered IAM systems offer significant advantages through their ability to process vast quantities of data and recognize patterns imperceptible to human analysts. Experimental evaluations comparing AI and human performance in security operations centers have documented that advanced machine learning systems can simultaneously process and correlate data from hundreds of distinct security sensors across enterprise environments, a scale fundamentally beyond human cognitive capacity [2]. These systems excel at continuously monitoring user behaviors across multiple dimensions—including login times, device characteristics, network attributes, and interaction patterns—to establish dynamic baseline profiles for individual users. When deviations from these profiles occur, AI can instantaneously calculate risk scores and trigger appropriate security responses, often before human analysts could even begin their assessment. Comparative benchmarking studies demonstrate that AI-powered anomaly detection identifies suspicious login patterns in seconds, compared to minutes for trained security analysts reviewing the same events [2].

Capability	AI Systems	Human Analysts
Data Processing	High-volume, simultaneous monitoring	Limited to manageable event volume
Response Time	Instantaneous assessment	Requires review time
Contextual Understanding	Limited organizational context	Superior business context understanding
Anomaly Detection	Identifies baseline deviations	May miss gradual pattern changes
Novel Scenarios	Struggles with unprecedented situations	Can apply judgment to new scenarios
Explainability	Often operates as "black box"	Can articulate decision rationale
Adaptability	Requires structured updates	Dynamic adjustment to new information

Table 1: AI vs. Human Security Analysis Capabilities [2]

Despite these capabilities, AI systems face notable limitations. False positives remain a persistent challenge, with cross-sectional analyses of enterprise deployments revealing varying false positive rates depending on implementation maturity and training data quality [2]. These erroneous flags can result in legitimate user activities being incorrectly identified as suspicious, potentially resulting in unnecessary access denials that impact operational efficiency. More concerning are algorithmic biases that may disproportionately impact certain user groups based on their behavioral patterns or work requirements. Detailed investigations into algorithmic fairness in security systems have documented that employees with non-standard work patterns experience authentication challenges more frequently than those following conventional schedules, with remote workers facing additional verification requirements at higher rates than office-based counterparts [4].

Furthermore, many AI systems operate as "black boxes," making their decision-making processes opaque and difficult to audit. Comprehensive surveys of security professionals reveal that only a minority express high confidence in their ability to explain the rationale behind AI-driven security decisions when questioned by compliance auditors or senior leadership [3]. This raises significant accountability concerns in security contexts where transparency is paramount, particularly as many organizations report struggling to demonstrate regulatory compliance for fully automated security decisions under frameworks such as GDPR, HIPAA, and PCI-DSS [3].

3. Decision Authority Framework: Balancing AI and Human Oversight

Establishing an effective decision authority framework requires organizations to carefully evaluate various factors that influence the appropriate level of AI autonomy versus human oversight. Low-risk, routine access scenarios—such as standard workday logins from known devices and networks—may be suitable candidates for full AI automation. Longitudinal studies of security automation efficacy indicate that organizations can safely delegate a significant proportion of routine authentication decisions to AI systems without increasing security risk profiles, resulting in documented labor cost reductions across security operations teams [1].

Conversely, high-risk scenarios—including access to critical systems, unusual access patterns, or authentication from unfamiliar locations—typically warrant human review despite initial AI assessment. Empirical evaluations of hybrid decision models demonstrate that human intervention in high-risk access scenarios reduces false positive rates while simultaneously improving threat detection accuracy compared to fully automated systems, particularly for novel attack vectors not present in training data [2]. The framework of security automation implemented across diverse organizational environments shows that structured human-AI collaboration delivers optimal outcomes when decision authority is contextually assigned based on comprehensive risk assessment [3].

A comprehensive decision authority matrix should consider multiple dimensions of access context. Extensive field studies of IAM deployments reveal that system criticality represents the most significant risk factor, with financial and intellectual property repositories requiring more rigorous verification than general productivity applications [1]. User privilege level constitutes another crucial dimension, with security logs demonstrating that privileged accounts accessing sensitive resources from new geographic locations trigger human review in mature security implementations to mitigate potential credential abuse [3].

Factor	AI Autonomy Appropriate	Human Oversight Required
System Criticality	Non-sensitive resources	Financial systems, IP repositories
User Privilege	Standard user accounts	Privileged/administrator access
Access Context	Known devices, standard hours	New devices, unusual times/locations
Behavioral Patterns	Consistent with baseline	Significant deviations
Business Context	Routine operations	Mergers, reorganizations, launches
Regulatory Requirements	Minimal compliance concerns	Healthcare, financial, critical infrastructure

Table 2: Decision Authority Framework [3]

Contextual factors requiring subjective interpretation present particular challenges for automated systems. Detailed analyses of access patterns show that behavioral deviations occurring during known business events (such as mergers, product launches, or quarterly financial processes) are more likely to represent legitimate activities than similar anomalies during routine operations, necessitating human judgment to properly contextualize [2]. Regulatory considerations introduce additional complexity, with healthcare organizations maintaining human oversight for a larger proportion of access decisions compared to less regulated industries due to specific compliance requirements and potential consequences of inappropriate data access [3].

Organizational security posture and risk tolerance significantly influence appropriate delegation levels. Quantitative assessment of security outcomes across enterprises demonstrates that organizations with formalized risk management frameworks implement graduated human-AI collaboration models that scale human involvement proportionally with risk levels, resulting in fewer security incidents while maintaining operational efficiency [4]. The technical capabilities of deployed AI systems represent the final critical factor, with advanced implementations employing ensemble learning methods demonstrating lower false positive rates than single-algorithm approaches, particularly when combining behavioral, contextual, and historical data points [2].

The most effective implementations adopt a sliding scale approach where AI autonomy increases for low-risk scenarios while ensuring human review remains integral for high-stakes decisions. Longitudinal performance data indicates that organizations implementing such flexible frameworks experience fewer security incidents while reducing overall IAM operational costs compared to organizations with static decision authority models [1]. As AI capabilities evolve, this balance requires continuous reassessment and calibration to maintain optimal security posture while respecting user experience considerations, with leading organizations adjusting their decision thresholds quarterly based on evolving threat intelligence and performance metrics [3].

4. Case Studies in Collaborative Security Models

Several organizations have successfully implemented collaborative human-AI security models that demonstrate the potential of thoughtfully designed decision authority frameworks. Experimental studies examining the integration of AI and human decision-making processes have shown that hybrid approaches consistently outperform either pure human or pure algorithmic decision strategies across multiple security domains. Research on artificial intelligence and human decision-making indicates that optimal security outcomes emerge when decision authority is dynamically allocated based on the comparative advantages of each intelligence type, with machines excelling at pattern recognition across large datasets while humans provide superior contextual reasoning in ambiguous scenarios [5].

A multinational financial institution's implementation of a tiered response system exemplifies this balanced approach. Comprehensive analysis of this case revealed that by creating clear delineation between routine and exceptional access scenarios, the organization achieved significant improvements in both security efficacy and operational efficiency. The sophistication of this implementation lies in its dynamic nature, with decision thresholds continuously adjusted based on real-time threat intelligence and historical performance data. Studies of AI and human collaboration in advanced cybersecurity applications have demonstrated that such contextually-adaptive frameworks consistently outperform static models, particularly in financial services environments where both attack sophistication and legitimate transaction volumes continue to increase exponentially [6].

In healthcare environments, collaborative models have demonstrated particular value given the unique tension between security requirements and clinical imperatives. A major healthcare provider implemented a contextually-aware IAM system specifically designed to account for the complex and often unpredictable nature of clinical workflows. Research examining cybersecurity implementations in healthcare settings has emphasized the critical importance of reconciling strict authentication protocols with clinical realities, particularly in emergency scenarios where traditional verification mechanisms may impede life-saving interventions. Analysis of this implementation revealed that by incorporating sophisticated contextual awareness into access decisions, the organization simultaneously strengthened security posture and improved clinical efficiency by reducing authentication-related delays during critical care situations [7].

These cases illustrate how effectively designed collaborative models can leverage the respective strengths of both AI and human decision-makers while mitigating their individual limitations. Longitudinal research on collaborative security frameworks has identified several critical success factors across implementations, including clear decision authority boundaries, transparent escalation pathways, continuous performance monitoring, and structured knowledge transfer between human and machine components. Studies of artificial intelligence and human decision interactions emphasize that successful security collaborations depend not only on technical implementation details but also on fundamental organizational factors including security governance models, risk tolerance frameworks, and cultural alignment around security objectives [5].

5. Ethical and Practical Considerations

The delegation of security decisions to AI systems raises significant ethical considerations beyond mere technical efficacy. Research examining the ethical dimensions of AI in cybersecurity has identified several fundamental tensions requiring thoughtful resolution as organizations increasingly rely on automated decision systems. Ethical analyses of AI in security contexts emphasize that privacy implications emerge when AI systems monitor increasingly granular aspects of user behavior to establish baseline patterns. Studies indicate that behavioral monitoring for security purposes creates inherent friction with privacy expectations, requiring organizations to carefully balance legitimate security requirements with reasonable employee privacy expectations [8].

Questions of fairness arise when algorithmic systems may inadvertently disadvantage users with atypical work patterns or responsibilities. Studies of AI and human collaboration in cybersecurity operations have documented that without specific attention to fairness considerations, automated security systems frequently generate disproportionate friction for certain user categories, including remote workers, employees with flexible schedules, and staff with specialized access requirements. Research indicates that this disparate impact often results from training data limitations rather than intentional discrimination, yet the operational consequences for affected users remain significant regardless of underlying causes [6].

Accountability concerns become prominent when determining responsibility for security incidents that occur despite—or because of—AI oversight. Ethical frameworks for AI in cyber security emphasize the fundamental challenge of establishing clear lines of responsibility in hybrid decision systems, particularly in regulatory environments that traditionally assume human decision authority. Research examining healthcare cybersecurity implementations has highlighted the particular complexity of accountability questions in clinical environments, where patient outcomes may be affected by security decisions yet traditional liability frameworks may not adequately address algorithmic contributions to decision processes [7].

From a practical perspective, explainability emerges as a critical requirement for sustainable implementation. Studies of human-AI collaboration in cybersecurity have demonstrated that security personnel consistently struggle to trust or effectively augment algorithmic decisions they cannot understand, regardless of demonstrated system accuracy. Research on the integration of artificial intelligence and human decision processes indicates that explainability requirements vary considerably by stakeholder type, with frontline security analysts requiring detailed technical explanations while executive decision-makers and affected users typically need conceptual clarity rather than algorithmic specificity [5].

Training and adaptation mechanisms represent another crucial consideration for maintaining system efficacy over time. Comprehensive studies of AI and human collaboration for advanced cybersecurity applications have documented the vulnerability of static models to both concept drift and adversarial manipulation. Research indicates that without structured refreshment processes, model accuracy typically degrades significantly within months as both legitimate user behaviors and attack methodologies evolve. Organizations implementing successful collaborative frameworks invariably incorporate formal mechanisms for continuous learning, typically combining automated adaptation with structured human review of model adjustments [6].

Appeal processes for users experiencing false positive restrictions have proven essential for maintaining organizational trust in security systems. Research in healthcare cybersecurity has emphasized the operational impact of false positives in clinical environments, where inappropriate access restrictions may directly affect patient care. Studies indicate that organizations implementing accessible and responsive remediation pathways achieve significantly higher user compliance with security protocols compared to those with opaque or cumbersome appeal mechanisms. This finding aligns with broader research on procedural justice in organizational contexts, suggesting that perceived fairness directly influences security cooperation independent of substantive outcomes [7].

Documentation and auditing procedures represent critical infrastructure for ensuring accountability in hybrid decision systems. Ethical analyses of AI in cybersecurity consistently emphasize the importance of comprehensive audit trails capturing both algorithmic recommendations and human oversight decisions. Research examining cybersecurity in healthcare settings has highlighted the particular importance of such documentation in regulated environments, where demonstrating appropriate governance of security decisions represents a fundamental compliance requirement. Studies indicate that organizations struggling with regulatory scrutiny most commonly identify documentation deficiencies rather than substantive security inadequacies as the primary compliance obstacle [7].

User education about AI-human collaboration in security processes emerges as a foundational requirement for effective implementation. Research on artificial intelligence and human decision making emphasizes that user understanding of security models directly influences compliance behaviors and threat reporting. Studies indicate that organizations effectively communicating the complementary roles of AI and human oversight achieve significantly higher levels of user cooperation with security protocols. This finding suggests that beyond technical implementation considerations, the narrative framing of collaborative security models significantly influences their practical effectiveness [5].

Organizations that neglect these ethical and practical dimensions risk implementing technically sophisticated systems that nevertheless undermine organizational trust and potentially introduce new vulnerabilities through unaddressed systemic weaknesses. Ethical frameworks for AI in cybersecurity emphasize that technical capability and ethical implementation represent distinct evaluation dimensions, with overall security effectiveness ultimately depending on both rather than either in isolation. Studies examining implementation outcomes consistently identify ethical and practical considerations as determinative factors in distinguishing successful from unsuccessful deployments, regardless of the technical sophistication of the underlying algorithms [8].

6. Future Directions and Recommendations

The evolution of human-AI collaboration in identity security will likely follow several trajectories as both technological capabilities and threat landscapes continue to develop. Research on artificial intelligence and human decision making suggests that future security frameworks will increasingly emphasize complementary capabilities rather than competitive performance between human and machine components. Studies examining emerging security paradigms identify predictive security models as a particularly promising direction, leveraging advanced analytics to anticipate potential threats before they materialize. This approach fundamentally shifts security postures from reactive to proactive stances, potentially preventing compromise attempts rather than merely detecting them after initiation [5].

Contextually-aware systems represent another significant advancement, incorporating broader organizational patterns and environmental factors into security assessments. Research on AI and human collaboration for advanced cybersecurity applications has demonstrated that contextual awareness significantly enhances detection accuracy while simultaneously reducing false positives compared to context-free analysis. Studies indicate that by incorporating factors like business cycles, organizational restructuring events, and external threat intelligence, these systems achieve more nuanced understanding of access legitimacy than traditional rule-based approaches. The most sophisticated implementations dynamically adjust security thresholds based on continuously updated contextual understanding, essentially modeling the organization's evolving risk landscape rather than applying static security rules [6].

Recommendation	Approach	Benefits
Incremental Implementation	Begin with low-risk scenarios	Reduced disruption, improved understanding
Comprehensive Metrics	Balance security, operational, user experience	Sustainable security with empirical feedback
Feedback Mechanisms	Structured analyst insights integration	Enhanced adaptation to emerging threats

Regular Reassessment	Formalized review cycles	Dynamic adaptation to evolving threats
Contextual Awareness	Business system integration	Nuanced understanding of access legitimacy
User-Centric Design	Security with user experience focus	Higher compliance, reduced friction

Table 3: Implementation Recommendations [6]

Self-healing security frameworks constitute a third promising direction, automatically adjusting to evolving attack vectors based on observed patterns. Studies examining cybersecurity in healthcare environments have documented the particular value of adaptive defense mechanisms in protecting critical infrastructure against rapidly evolving threats. Research indicates that self-healing approaches demonstrate substantial resilience against novel attack methodologies, particularly when combining unsupervised anomaly detection with traditional signature-based protections. These systems leverage machine learning techniques to identify emerging attack patterns and automatically implement appropriate countermeasures, often without requiring explicit human intervention for known threat categories [7].

Organizations seeking to optimize their identity security frameworks should consider several evidence-based recommendations emerging from the research literature. Studies on the integration of artificial intelligence and human decision making consistently advocate for incremental implementation approaches, beginning with low-risk scenarios while maintaining comprehensive human oversight for critical decisions. Research indicates that organizations adopting phased deployment strategies experience fewer security disruptions during transition periods while simultaneously developing more nuanced understanding of appropriate decision authority boundaries through practical experience. This incremental approach not only mitigates potential security risks during the transition period but also allows for organizational learning and adjustment before expanding automation to more sensitive contexts [5].

Establishing clear metrics for both security efficacy and user experience impacts emerges as another critical recommendation. Research on AI and human collaboration for advanced cybersecurity applications emphasizes the importance of comprehensive measurement frameworks that capture the multidimensional nature of security outcomes. Studies indicate that organizations implementing balanced metrics spanning traditional security indicators, operational efficiency measures, and user experience dimensions achieve more sustainable security postures than those focusing exclusively on technical security metrics. These balanced frameworks provide essential feedback for continuously refining decision authority boundaries between human and machine components based on empirical performance data rather than theoretical assumptions [6].

Developing robust feedback mechanisms between human analysts and AI systems represents a fundamental requirement for long-term efficacy. Studies examining cybersecurity in healthcare environments have demonstrated that structured knowledge transfer between human experts and machine learning systems significantly enhances adaptation to emerging threats. Research indicates that organizations implementing formal processes for incorporating analyst insights into algorithmic models achieve substantial performance improvements compared to those relying exclusively on automated learning. These bidirectional feedback loops prove particularly valuable when addressing novel attack vectors not represented in initial training data, essentially allowing the system to benefit from human contextual understanding while maintaining algorithmic efficiency [7].

Implementation of transparent processes for documenting decision rationales represents another consistent recommendation from the research literature. Ethical frameworks for AI in cybersecurity emphasize the importance of explainable security decisions, particularly for access denials that significantly impact user productivity. Studies indicate that understanding the reasoning behind security decisions substantially increases user acceptance and cooperation, even when those decisions create operational friction. This transparency becomes particularly important when addressing false positives, where clear explanation can help maintain organizational trust despite occasional system limitations [8].

Regular reassessment of the decision authority framework emerges as essential as technologies mature and organizational needs evolve. Research on artificial intelligence and human decision making emphasizes the dynamic nature of optimal collaboration models, with ideal authority boundaries shifting as both technological capabilities and threat landscapes evolve. Studies examining implementation outcomes consistently identify framework adaptation as a key differentiator between successful and unsuccessful deployments. Research on cybersecurity in healthcare indicates that leading organizations formalize this reassessment process, reviewing and adjusting decision thresholds on recurring cycles based on performance metrics, emerging threats, and evolving business requirements [7].

The future of identity security does not lie in AI replacing human judgment, but rather in thoughtfully designed collaborative models that leverage the unique strengths of both. Comprehensive analyses of security outcomes across diverse organizational contexts consistently demonstrate that integrating algorithmic efficiency with human contextual understanding yields superior results compared to either approach in isolation. Ethical frameworks for AI in cybersecurity emphasize that maximizing security effectiveness ultimately depends on appropriate integration rather than wholesale delegation or retention of decision authority. By carefully considering when AI should decide and when human expertise remains essential, organizations can develop identity security frameworks that are simultaneously more robust and more responsive to legitimate user needs [8].

7. Conclusion

The evolving landscape of identity security necessitates thoughtful integration of artificial intelligence capabilities with human expertise, rather than wholesale delegation or retention of decision authority. Research demonstrates that collaborative security frameworks consistently outperform both fully automated and exclusively human-driven approaches across diverse organizational contexts. This performance advantage stems from the complementary nature of algorithmic and human intelligence: machines excel at rapidly processing vast datasets and identifying patterns, while humans provide superior contextual understanding and ethical judgment in ambiguous scenarios. Successful implementations share common characteristics, including clear decision authority boundaries, contextually-adaptive response frameworks, transparent escalation pathways, and structured knowledge transfer mechanisms. Organizations that thoughtfully design these collaborative frameworks strengthen security posture while improving operational efficiency and user experience. The ethical dimensions prove equally important as technical details, with privacy considerations, fairness mechanisms, and accountability structures representing critical success factors. Looking forward, human-AI collaboration will likely emphasize predictive capabilities, contextual awareness, and self-healing frameworks. The future of identity security lies not in AI replacing human judgment, but in collaborative models that leverage the unique strengths of both.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Irshaad Jada, et al, "The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review," Data and Information Management, Volume 8, Issue 2, June 2024, Available: https://www.sciencedirect.com/science/article/pii/S2543925123000372
[2] Shivani Bhinge, et al, "Quantifying the Performance Gap between Real and Ai-Generated Synthetic Images in Computer Vision," January 2023, Online, Available: https://www.researchgate.net/publication/374512207_Quantifying_the_Performance_Gap_between_Real_and_Ai-Generated_Synthetic_Images_in_Computer_Vision
[3] M.Robinson Joel, et al, "Framework of Smarthome Automation and Security," April 2023, International Journal of Scientific Research in Science Engineering and Technology, Available: https://www.researchgate.net/publication/370516218_Framework_of_Smarthome_Automation_and_Security
[4] Anjali Sandeep Gaikwad, "ALGORITHMIC BIAS DETECTION AND MITIGATION: ARTIFICIAL INTELLIGENCE," September 2022, Online, Available: https://www.researchgate.net/publication/363248756_ALGORITHMIC_BIAS_DETECTION_AND_MITIGATION_ARTIFICIAL_INTELLIGENCE
[5] Jean-Charles Pomerol, "Artificial intelligence and human decision making," March 1997, European Journal of Operational Research, Available: https://www.researchgate.net/publication/222506083_Artificial_intelligence_and_human_decision_making
[6] Kapil Manshani, Iaeme Pub, "AI AND HUMAN COLLABORATION FOR ADVANCED CYBERSECURITY: REAL-TIME THREAT DETECTION AND RESPONSE," February 2025, Available: https://www.researchgate.net/publication/389599486_AI_AND_HUMAN_COLLABORATION_FOR_ADVANCED_CYBERSECURITY_REAL-TIME_THREAT_DETECTION_AND_RESPONSE
[7] Yash Patel, "Cybersecurity in Healthcare: Protecting Critical Infrastructure Against Evolving Threats," November 2024, International Journal of Computer Trends and Technology, Available: https://www.researchgate.net/publication/386348803_Cybersecurity_in_Healthcare_Protecting_Critical_Infrastructure_Against_Evolving_Threats
[8] Emmanuel Chris, "Ethical Considerations in AI for Cyber Security," December 2022, Online, Available: https://www.researchgate.net/publication/387958291_Ethical_Considerations_in_AI_for_Cyber_Security