Communication in Physical Sciences, 2026, 13(2): 403-421

Human-AI Collaboration in Cybersecurity Decision-making: A Systematic Review of Literature

Dahunsi Samuel Adeyemi.

Received: 23 January 2026/Accepted: 14 March 2026 /Published: 20 March 2026
https://dx.doi.org/10.4314/cps.v13i3.5

Abstract: *Artificial intelligence has become an important tool in cybersecurity decision-making, but there are arguments about balancing it with human cognition. Human cognition can significantly influence cybersecurity outcomes, making it essential to examine human–AI interaction within cybersecurity environments. Thus, this study examined human-AI collaboration in cybersecurity decision-making using systematic review of literature. Using the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA), a total of seventeen (17) articles were selected from credible databases using some set of inclusion and exclusion criteria. The study findings showed that human-AI collaboration in cybersecurity decision-making is a complementary and symbiotic approach wherein AI enhances human judgement through structured frameworks and architectures. Results showed that AI performs real-time monitoring and analysis while humans handle complex or high-risk decision, supported by dynamic role-based models that allow flex collaboration between human and AI. Findings indicate that collaborative human-AI cybersecurity decision-making is supported by a combination of technical and organizational techniques, which enhance transparency, trust, accuracy, and alignment between human judgement and AI outputs. Findings showed that as a result of its efficiency, scalability, and adaptability, human-AI collaboration in cybersecurity decision-making is more effective than human-only or AI-only approaches. The study concludes that despite the challenges of using human-AI collaboration in cybersecurity decision-making, it is more effective than using solo approach – whether human or AI only.*

Keywords: *Cybersecurity, cybersecurity decision-making, AI cybersecurity, human cybersecurity, human-AI collaboration*

Dahunsi Samuel Adeyemi.
Department of Computer Science and Cybersecurity, University of Central Missouri Missouri, USA.
College of Health, Science and Technology,.
Email: dxa26930@ucmo.edu
https://orcid.org/0009-0007-5485-8052

1.0 Introduction

Cybersecurity decision-making refers to the processes through which individuals and organizations evaluate cyber risks and select appropriate strategies, policies, and technical responses to mitigate threats. It has been established that cybersecurity decision-making involves the identification and operationalization of a tailored approach to address risk management and cybersecurity problems. This helps to improve the ability of company leaders to handle cyber threats (Goel *et al*., 2020). Usually, decision-making with respect to cybersecurity is linear, starting from the individual to the larger organizations/institutions. For instance, Al-Hashem & *Et al* (2023) noted that cybersecurity decision-making extends from individuals to organizations, where organizational leaders make important choices regarding budgets, policies, and incident response. This explains the importance of adopting a multidisciplinary approach to support cybersecurity decision-making, especially a combination of computer science and economics. Organizations can adopt decision techniques and economic indicators to evaluate cybersecurity investments (Beissel, 2016). Despite these advances, cybersecurity decision-making remains complex due to increasing threat sophistication, uncertainty, and the limitations of purely technical or purely human-centered approaches.

Previous studies demonstrate that decision-support and recognition systems can improve cybersecurity outcomes; however, existing frameworks still struggle to simultaneously address threat, vulnerability, and consequence dimensions of risk assessment (Akhmetov *et al*., 2017; Ganin *et al*., 2020).

This underscores the importance of decision-analysis-based approach to assess the overall utility of cybersecurity management alternatives. This helps to evaluate and rank different cybersecurity enhancement strategies, which may not be achievable without proper cybersecurity decision-making.

Meanwhile, cybersecurity decision-making is essential for reducing uncertainty and preventing inconsistent security practices. Fielder *et al*. (2018) established that there are uncertainties in conducting cybersecurity risk assessment, which often affects cybersecurity investments and resulting optimal strategies. They proposed different models that support security managers with decisions regarding the optimal allocation of financial resources under uncertainty. In a comparison of performance of experienced professionals against an inexperienced control group, Jalali *et al*. (2019) noted that experienced subjects were better at learning the need for proactive decision-making through iteration, though neither group understood delay mechanisms well. The study highlights the importance of training decision-makers with a focus on systems thinking skills to improve cybersecurity decision-making. This leaves opportunity for driving cybersecurity decision-making through human-AI collaboration. These limitations highlight the need for decision frameworks that combine human reasoning capabilities with advanced computational intelligence.

There are several approaches to achieve cybersecurity decision-making, which may rely on either artificial intelligence (AI), human expertise, or a combination of both. Jimmy (2021) noted that AI can analyze vast amounts of data, detect patterns, and identify anomalies at speeds and scales beyond human capabilities. This enables real-time detection and response to threats in cybersecurity. In fact, the author discussed how explainable AI (XAI) can provide insights into the decision-making process of AI algorithms, which help cybersecurity professionals to understand and interpret system outputs for more effective threat response. These capabilities support more informed and timely cybersecurity decision-making within organizational environments.

Abdulhussein (2024) addressed the decision-making issues in AI technology investments and cybersecurity, emphasizing collaboration, awareness, financial factors, and adaptability to evolving threats.

Furthermore, Mahadik *et al*. (2024) adduced that using the hybrid method of rule-based system with the Random Forest machine learning algorithm, which is an element of AI, improve the accuracy and efficiency of cybersecurity decision-making. This underscores AI effectiveness in real-world applications to conduct cybersecurity decision-making(Ajiboye *et al*., 2025; Samakinde & Arohunmolase, 2026). Despite these advantages, several challenges remain associated with fully automated AI-driven cybersecurity decision-making. Bhardwaj and Choudhary (2024) noted that the use of AI-driven tools has transformed the approach to cybersecurity with respect to the mitigation of cyber risks in real-time. However, with the applications of these systems, there are still challenges associated with intrusion detection, fraud prevention, malware analysis, and secure network management. This accentuates a more effective way of ensuring cybersecurity decision-making through AI-human collaboration. Consequently, integrating human expertise with AI capabilities has emerged as a promising approach to enhance the reliability and accountability of decisions.

While artificial intelligence-driven tools have been seen as important in cybersecurity decision-making, several arguments have been made to ensure that these AI-driven tools are balanced with human cognition. Cognitive biases within human decision-making processes can significantly influence cybersecurity outcomes, which makes it important to consider nuances in the assessment of human factors in protection policies (Al, 2024). This has intensified scholarly debate regarding the development of cybersecurity systems that effectively integrate human cognitive processes.

.Hagen *et al*. (2025) noted that cognitive biases, automation, and confirmation bias have an influence on security analysts' trust in AI-driven tools, which ultimately impacts cybersecurity decision-making. Al-Hashem and *Et al* (2023) added that, aside from cognitive biases, risk perception and the availability of information often occurring under pressure affect decision-making in the face of cyber threats. This necessitates the need for human factors in AI-driven cybersecurity.

Over time, human cognition in cybersecurity decision-making is important. It serves an oversight function to identify and correct any possible omission or bug that may arise through algorithmic design or framework (Adeyemi, 2023). Ceric and Holland (2019) suggested that managers' decision-making processes regarding cybersecurity are often influenced by heuristics and inherent cognitive biases, which can increase organizational risk. Kritika (2026) noted that there are new tools and perspectives for studying the human factor in cybersecurity, which involves bridging economics, psychology, and neuroscience to provide biological insights into human decision-making and behavior. The human mind provides light on cognitive complexities, which shows the complex interaction between the human mind and the virtual world (Al-Hashem *et al*, 2023).

However, despite increasing recognition of both AI capabilities and human cognitive contributions, existing research remains fragmented, with limited systematic synthesis examining how human–AI collaboration supports cybersecurity decision-making processes. This buttresses the importance of the human mind in risk perception, drawing on psychological theories and empirical evidence. Therefore, this study systematically reviews existing literature to examine how human–AI collaboration is conceptualized, implemented, and evaluated in cybersecurity decision-making contexts. This study will provide answers to questions, which are itemized as follows:

1. How is human-AI collaboration conceptualized and implemented in cybersecurity decision-making?
2. What are the human-AI collaborative approaches employed in cybersecurity decision-making tasks?
3. What are the techniques used to support collaborative human-AI cybersecurity decision-making?
4. What are the impacts of human-AI collaboration compared to human-only or an AI-only approach?
5. What are the challenges identified regarding human-AI collaboration in cybersecurity decision-making?

2.0 Methodology

This study adopts a qualitative systematic review design to examine human–AI collaboration in cybersecurity decision-making. The design follows a structured evidence-synthesis approach that systematically evaluates existing empirical studies on human–AI collaboration in cybersecurity decision-making.

The systematic review research design allows for the formulation of research questions, search for relevant literature from credible databases, download literature that meet the set inclusion criteria, assess the qualities of the final selected literature, extract needed information from the final selected literature/studies, and analyse the information collected from the final selected literature to generate themes (Schut *et al*., 2024). This structured approach enhances methodological rigor, transparency, and reproducibility, thereby ensuring scientific reliability of the review process (Adeyemi *et al*., 2025).

The systematic review process involved formulating research questions, identifying relevant studies from selected databases, applying predefined inclusion and exclusion criteria, assessing study quality, extracting relevant data, and synthesizing findings to generate thematic insights (Schut *et al*., 2024). To address the study's research questions, a structured and transparent review protocol was implemented to enhance repeatability and credibility. This involved the use of predefined search terms and systematic search strategies to identify relevant literature. The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

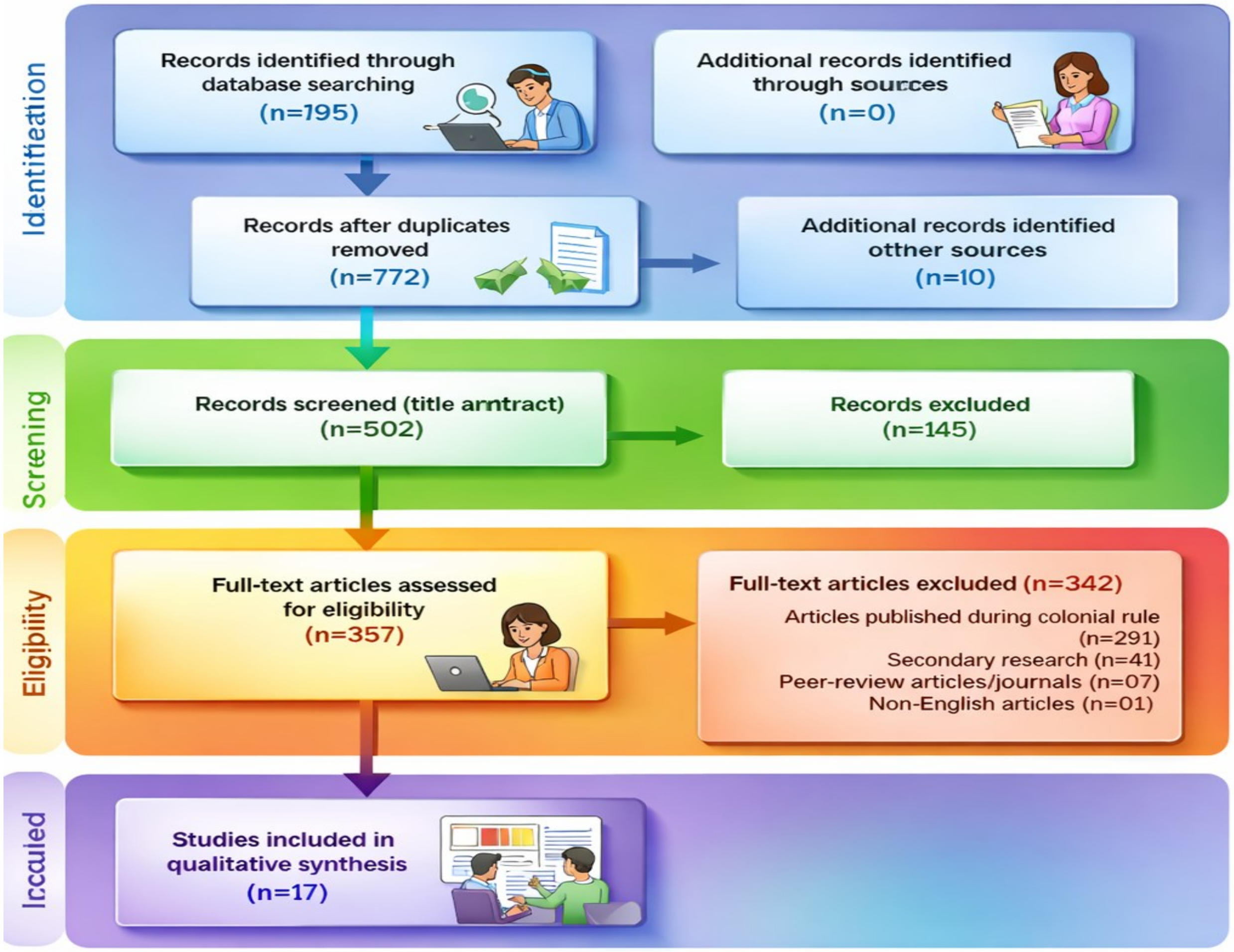
Identifieation
Records identified through database searching (n=195)
Additional records identified through sources (n=0)
Records after duplicates removed (n=772)
Additional records identified otther sources (n=10)
Screening
Records screened (title arntract) (n=502)
Records excluded (n=145)
Eligibiity
Full-text articles assessed for eligibility (n=357)
Full-text articles excluded (n=342)
Articles published during colonial rule (n=291)
Secondary research (n=41)
Peer-review articles/journals (n=07)
Non-English articles (n=01)
Iccuied
Studies included in qualitative synthesis (n=17)

Fig. 1: PRISMA diagram flow (Author's self-designed, 2025)

This is a framework designed to ensure structured and credible data collection in systematic reviews (Adeyemi *et al*., 2025). It is widely used for systematic reviews of literature; hence, it was considered for this study. The framework has 27 items, which are categorized into identification, screening, eligibility, and inclusion (Helach *et al*., 2023). PRISMA consists of four main stages: identification, screening, eligibility, and inclusion (Helach *et al*., 2023). The identification stage involves database searching and record collection; screening evaluates titles and abstracts; eligibility assesses full-text articles using predefined criteria; and inclusion determines the final studies used for synthesis. Using the PRISMA procedure, seventeen (17) studies were ultimately selected for qualitative synthesis and data extraction.

Five academic databases were searched to identify relevant literature on human–AI collaboration in cybersecurity decision-making.

These databases include Scopus, Google Scholar, Taylor and Francis, Emerald, and Web of Science. These databases were considered due to their extensive coverage of peer-reviewed research relevant to the study topic. on the topic of discourse. To retrieve optimum and relevant information, different search terms were used for the literature search, which are core to the major aim of this study and the specific research questions identified. Some of the search terms used include "conceptualization and implementation of human-AI collaboration in cybersecurity decision-making", "human-AI collaborative approaches employed in cybersecurity decision-making tasks", "techniques used to support collaborative human-AI cybersecurity decision-making", "impacts of human-AI collaboration compared to human-only or AI-only approach", and "challenges associated with human-AI collaboration in cybersecurity decision-making".Boolean operators (e.g., "OR") were applied to broaden search coverage and capture related terminology.

No temporal restriction was applied during the search process in order to capture the full evolution of research on human–AI collaboration in cybersecurity decision-making.

Table 1 presents the electronic search strategy adopted for this study, including the search terms applied across the selected databases, the number of retrieved records, and the sequential screening process that led to the final selection of studies for qualitative synthesis.

All the collected or extracted evidence from the final selected literature were analyzed using the "a priori" thematic analysis, which involves using some predetermined themes to analyze the data (Adeyemi *et al*., 2025).

Table 1: Electronic Search Strategy (Extracts for five databases)

S/N	**Search terms**	**Web of Science**	**Scopus**	**Google Scholar**	**Emerald**	**Taylor and Francis**
				Number of hits		
S1	Conceptualization and implementation of human-AI collaboration in cybersecurity decision-making	5821	4732	1293	2405	4095
S2	Human-AI collaborative approaches employed in	4532	5320	1384	1930	3281

	cybersecurity decision-making tasks					
S3	Techniques used to support collaborative human-AI cybersecurity decision-making	4000	3000	7000	2300	5200
S4	Impacts of human-AI collaboration compared to human-only or AI-only approach	7100	15000	21000	1800	4300
S5	Challenges associated with human-AI collaboration in cybersecurity decision-making	19000	1500	5000	1540	3100
S6	Human-enhanced cybersecurity decision-making OR AI-enhanced cybersecurity decision-making	1500	800	750	1260	2150
Databases search limits adopted						
Duplicates removed		107	113	181	85	95
Titles and abstracts checked		83	88	120	45	71
Secondary research		50	54	83	40	38
Peer-reviewed articles/journals		07	05	07	05	02
English language only		NA	N/A	02	N/A	N/A
Final selected		4	3	5	5	0

Source: Author's Literature Search (2025)

3.0 Results and Discussion

This section presents and discusses the findings of the systematic review according to the study's research questions, synthesizing evidence across the selected studies to identify dominant themes and patterns in human–AI collaboration in cybersecurity decision-making. For the research question one, which is based on the conceptualization and implementation of human-AI collaboration in cybersecurity decision-making, the findings indicate that human and AI systems play complementary roles. . The review demonstrates that human–AI collaboration is predominantly conceptualized as a symbiotic or augmented intelligence paradigm, in which AI systems enhance rather than replace human judgement. Three of the final selected studies (Malatji, 2025; Saadallah *et al*., 2025; Van Hoang, 2023) frame collaboration as a structured integration of human judgement, ethical reasoning, and contextual awareness with AI-driven analytics and automation. The implementation occurs through formal frameworks and architectural models, which include symbiotic integration models (Saadallah *et al*., 2025), collaborative intelligence frameworks (Van Hoang, 2023), and augmented intelligence architectures (Malatji, 2025). Within operational environments such as Security Operations Centers (SOCs), collaboration is operationalized through agentic AI systems that embed human oversight mechanisms within automated cybersecurity workflows.

(Yaich *et al*., 2025). This conceptualization reflects a broader shift in cybersecurity governance from automation-centred models toward human-centred intelligent systems that prioritize accountability and contextual reasoning.

With respect to Research Question 2, which examines human–AI collaborative approaches in cybersecurity decision-making tasks. The results showed that there are varying degrees of AI autonomy and human control, which include Humans-in-the-Loop (HITL), Humans-on-the-Loop (HOTL), Humans-in-Command (HIC), Humans-alongside-the-Loop (HATL), and coactive systems (Malatji, 2024). These models represent varying levels of human authority and AI autonomy, illustrating how decision responsibility is dynamically distributed between human analysts and intelligent systems.

Two of the final selected studies (Karunamurthy *et al*., 2023; Van Hoang, 2023) highlight that AI systems are responsible for real-time monitoring, anomaly detection, intrusion detection, and large-scale data analysis, while humans intervene in ambiguous, high-risk, or ethically sensitive decisions. Results indicate that there are advanced collaborative approaches, which introduced dynamic role-based agent models such as assistant, autopilot, companion, and operator roles (Yaich *et al*., 2025). All of these allow collaboration modes to shift across the cybersecurity lifecycles. The presence of multiple collaboration modes suggests that effective cybersecurity decision-making depends on adaptive interaction rather than fixed human or machine dominance.

On the research question three, which focuses on techniques supporting collaborative human-AI cybersecurity decision-making, the findings identify a range of technical and organizational techniques employed to support effective collaboration. Three of the final selected studies (Desai *et al*., 2024; Safavi *et al*., 2025; Wairagade & Ranjan, 2025) highlight Explainable Artificial Intelligence (XAI) as the most frequently reported enabling mechanism amongst all the techniques, with methods such as SHAP, LIME, Explainable Boosting Machines, and Grad-CAM to improve transparency, trust, and interpretability in AI-driven cybersecurity decisions. Other techniques used include iterative feedback loops, continuous model validation, and a reliability scoring mechanism to align AI outputs with human expectations (Van Hoang, 2023). The parallel and multi-channel decision architectures were found to facilitate human and machine reasoning with the benefit of accuracy and responsiveness (Dolgikh & Mulesa, 2021). Two of the final selected studies (Blasch *et al*., 2023; Malatji, 2025) highlighted that risk-aware tools (e.g., decision matrices, uncertainty bounds, and multi-attribute scorecard tables) are used to support shared decision-making and certification of AI systems. Collectively, these techniques reduce the opacity of AI systems and strengthen human trust, thereby improving collaborative decision reliability in high-risk cybersecurity environments.

On research question four, which focuses on the impacts of human-AI collaboration compared to human-only or AI-only approaches, results showed that human-AI collaboration approach is more impactful than the AI-only and human-only approaches in cybersecurity decision-making. The collaborative approach was found to be more accurate, efficient, scalable, and adaptable to evolving threats(Alam & Khan, 2024; Dolgikh & Mulesa, 2021; Gyles & Boonthum-Denecke, 2025)

The synthesis indicates that human–AI collaborative approaches outperform both human-only and AI-only models across multiple performance dimensions, including accuracy, efficiency, scalability, and adaptability. Gyles & Boonthum-Denecke, 2025). Additionally, Alam and Khan (2024) showed that using AI-human collaborative approach enhances user satisfaction compared to existing mechanisms. These findings support the emerging consensus that hybrid intelligence systems mitigate the limitations inherent in purely automated or purely human-driven cybersecurity decision processes.

On Regarding Research Question 5, which examines challenges associated with human–AI collaboration in cybersecurity decision-making, results showed that despite its advantages, there are several challenges associated with human-AI collaboration in cybersecurity decision-making. The most frequently reported challenges include trust and explainability, with cybersecurity experts expressing scepticism towards opaque AI outputs and high false-positive rates (Hagen *et al*., 2025; West & Hale, 2025). Additional challenges include cognitive biases, specifically automation bias and confirmation bias, which influences how analysts interpret and rely on AI recommendations (Hagen *et al*., 2025). Other challenges that were identified include privacy concerns, ethical accountability, model reliability, adversarial manipulations, and high implementation costs (Safavi *et al*., 2025; West & Hale, 2025). The reviewed studies suggest that these challenges can be mitigated through the adoption of explainable AI techniques, bias-awareness training, adaptive trust calibration mechanisms, and governance frameworks that maintain human oversight over critical cybersecurity decisions (Karunamurthy *et al*., 2023; Wairagade & Ranjan, 2025).

The findings reveal a converging trend toward hybrid intelligence in cybersecurity decision-making, where human expertise and artificial intelligence capabilities function as mutually reinforcing components. The reviewed literature consistently emphasizes that successful collaboration depends not only on technological advancement but also on organizational design, trust calibration, and ethical governance structures. This synthesis highlights the importance of integrating human cognitive strengths with AI analytical power to achieve resilient and adaptive cybersecurity systems.

4.0 Conclusion

This study demonstrates that human–AI collaborative approaches provide a more effective cybersecurity decision-making model than either human-only or AI-only systems. The findings position human–AI collaboration as a symbiotic paradigm in which artificial intelligence augments human judgment rather than replacing it. Implementation of this collaborative approach occurs through structured frameworks, architectural models, and agentic systems that embed human oversight within cybersecurity workflows, highlighting that optimal decision-making requires the integration of computational efficiency with human contextual awareness, ethical reasoning, and strategic judgement. Collaborative effectiveness is supported by mechanisms that strengthen transparency, trust, and shared situational awareness between human analysts and AI systems, including iterative feedback loops, validation processes, explainable artificial intelligence (XAI) techniques, and risk-aware decision tools. Despite challenges related to trust deficits, explainability limitations, cognitive biases, and ethical accountability, human–AI collaboration enhances decision accuracy, operational efficiency, scalability, and adaptability to evolving cyber threats, reflecting a sustainable and forward-looking approach for cybersecurity governance.

The findings carry significant implications across policy, practice, theory, and society. Policymakers and cybersecurity agencies are encouraged to develop regulatory and governance frameworks that define accountability, ethical responsibilities, and oversight functions in AI-enabled cybersecurity systems. Such frameworks should mandate explainability, auditability, and human control to mitigate risks associated with opaque algorithms and adversarial manipulation. Institutional adoption of human–AI collaboration can foster responsible AI deployment, ensuring that humans maintain oversight over consequential security decisions. For cybersecurity professionals, the study highlights the importance of adopting collaborative architectures that integrate human expertise with AI-driven analytics rather than relying solely on automated systems. The prevalence of XAI techniques underscores the need for transparency and interpretability to enhance analyst trust and operational usability, while continuous workforce development programs can address cognitive bias, improve AI literacy, and strengthen the ability of analysts to critically interpret AI outputs.

Theoretically, the study advances human–AI collaboration theory by reinforcing the augmented intelligence paradigm as a dominant conceptual framework within cybersecurity contexts. Future theoretical discourse should focus on measurable constructs for collaboration effectiveness, trust dynamics, and human–AI team performance in complex decision environments. It should also consider multiple collaboration modes, ranging from human-in-the-loop models to fully coactive systems, to account for dynamic role allocation, trust calibration, and cognitive interaction between humans and intelligent systems. From a societal perspective, effective human–AI collaboration enhances resilience against cyber threats, supporting the protection of critical infrastructure, financial systems, healthcare services, and personal data, thereby contributing to national security, economic stability, and public trust in digital systems.

5.0 References

Abdulhussein, M. (2024). *The impact of artificial intelligence and machine learning on organizations cybersecurity*. Liberty University.

Adeyemi, D. S. (2023). Autonomous response systems in cybersecurity: A systematic review of AI-driven automation tools. *Communication in Physical Sciences*, *9*(4), 878-898.

Adeyemi, I. O., Akanbi, M. L., & Issa, A. O. (2025). Influence of game literacy on gamified library services: A systematic review of literature. *Alexandria,* 1-15. https://doi.org/10.1177/09557490251335953

Ajiboye, A. A., Gaffari, M. A., Obamuwagun, O. E. (2025). Predictive Analytics in Sport Management: Applying Machine Learning Models for Talent Identification and Team Performance Forecasting. *Communication in Physical Sciences.* 12(7):2032-2048 ***https://dx.doi.org/10.4314/cps.v12i7.5***

Akhmetov, B., Lakhno, V., Boiko, Y., & Mishchenko, A. (2017). Designing a decision support system for the weakly formalized problems in the provision of cybersecurity. *Eastern-European Journal of Enterprise Technologies*, *1*(2), 85-96.

Al, Q. A. (2024). Human factors in cyber defense. *The Art of Cyber Defence: From Risk Assessment to Threat Intelligence*, 260.

Al-Hashem, N., et al. (2023). The psychological aspect of cybersecurity: Understanding cyber threat perception and decision-making. *International Journal of Applied Machine Learning and Computational Intelligence*, *13*(8), 11-22.

Alam, S., & Khan, M. F. (2024). Enhancing AI-human collaborative decision-making in Industry 4.0 management practices. *IEEE Access*.

Beissel, S. (2016). *Cybersecurity investments: Decision support under economic aspects*. Springer.

Bhardwaj, A., & Choudhary, S. K. (2024). AI Based Decision Support System for Cyber Forensics Investigations. In *2024 IEEE 4th International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1-9). IEEE.

Blasch, E., Bastian, N. D., Aved, A., & Ardiles-Cruz, E. (2023). Human-machine cooperative AI decision-making with heterogeneous data. In *Signal Processing, Sensor/Information Fusion, and Target Recognition XXXII* (Vol. 12547, pp. 162-171). SPIE.

Ceric, A., & Holland, P. (2019). The role of cognitive biases in anticipating and responding to cyberattacks. *Information Technology & People*, *32*(1), 171-188.

Desai, B., Patil, K., Mehta, I., & Patil, A. (2024). Explainable AI in Cybersecurity: A Comprehensive Framework for enhancing transparency, trust, and Human-AI Collaboration. In *2024 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 135-150). IEEE.

Dolgikh, S., & Mulesa, O. (2021). Collaborative human-AI decision-making systems. In *IntSol Workshops* (pp. 96-105).

Etuk, E. A., & Omankwu, O. C. B. (2025). Human-AI collaboration: Enhancing decision-making in critical sectors. *Communication In Physical Sciences*, *12*(2), 426-433.

Fielder, A., König, S., Panaousis, E., Schauer, S., & Rass, S. (2018). Risk assessment uncertainties in cybersecurity investments. *Games*, *9*(2), 1-34.

Ganin, A. A., Quach, P., Panwar, M., Collier, Z. A., Keisler, J. M., Marchese, D., & Linkov, I. (2020). Multicriteria decision framework for cybersecurity risk assessment and management. *Risk Analysis*, *40*(1), 183-199.

Goel, R., Kumar, A., & Haddow, J. (2020). PRISM: a strategic decision framework for cybersecurity risk assessment. *Information & Computer Security*, *28*(4), 591-625.

Gyles, S., & Boonthum-Denecke, C. (2025). AI-driven cybersecurity: Opportunities, challenges, and the future of human-AI collaboration. The 2025 ADMI Symposium.

Hagen, R. A., Øverlier, L., & Helkala, K. (2025). Human factors in AI-driven cybersecurity: Cognitive biases and trust issues. *Digital Threats: Research and Practice*, *6*(4), 1-20.

Helach, J., Hoffmann, F., Pieper, D., & Allers, K. (2023). Reporting according to the preferred reporting items for systematic reviews and meta-analyses for abstracts (PRISMA-A) depends on abstract length. *Journal of Clinical Epidemiology*, *154*, 167-177.

Jalali, M. S., Siegel, M., & Madnick, S. (2019). Decision-making and biases in cybersecurity capability development: Evidence from a simulation game experiment. *The Journal of Strategic Information Systems*, *28*(1), 66-82.

Jimmy, F. (2021). Emerging threats: The latest cybersecurity risks and the role of artificial intelligence in enhancing cybersecurity defenses. *Valley International Journal Digital Library*, *1*, 564-74.

Karunamurthy, A., Kiruthivasan, R., & Gauthamkrishna, S. (2023). Human-in-the-loop intelligence: Advancing AI-centric cybersecurity for the future. *Quing: International Journal of Multidisciplinary Scientific Research and Development*, *2*(3), 20-43.

Kritika, M. (2026). Neuro-cognitive approaches to cybersecurity: a systematic review integrating neuroscience and cognitive psychology for human factor analysis. *Information & Computer Security*, 1-29.

Maennel, K., & Maennel, O. M. (2024). Human-AI collaboration and cyber security training: Learning analytics opportunities and challenges. In *2024 17th International Conference on Security of Information and Networks (SIN)* (pp. 01-08). IEEE.

Mahadik, R. V., Kingsly Jabakumar, A., Dari, S. S., Mishra, S., Katikar, S. M., & Raghunath, M. P. (2024). Decision Support Systems Enhanced by Machine Learning in Cybersecurity. In *International Conference on Smart Computing and Informatics* (pp. 379-389). Singapore: Springer Nature Singapore.

Malatji, M. (2024). Evaluating human-machine interaction paradigms for effective human-artificial intelligence collaboration in cybersecurity. In *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)* (pp. 1268-1272). IEEE.

Malatji, M. (2025). Augmented intelligence framework for human–artificial intelligence teaming in cybersecurity. *Human-Centric Intelligent Systems*, 1-30.

Saadallah, M., Shahim, A., & Khapova, S. (2025). Optimizing ai and human expertise integration in cybersecurity: Enhancing operational efficiency and collaborative decision-making. *PriMera Scientific Engineering*, *6*(2), 03-20.

Safavi, S., Abdulnabi, M. S. H., Rana, M. E., & Alizadeh, S. (2025). From black box to trustworthy AI: A secure framework for explainable cybersecurity decision-making. In *2025 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)* (pp. 1-4). IEEE.

Samakinde, A. S., Arohunmolase, V. B. (2026). A Review of Machine Learning-Based Geochemical Signature Analysis for Mineral Prospectivity Mapping. Communication in Physical Sciences, 13(1): 36-59. https://dx.doi.org/10.4314/cps.v13i1.4

Schut, M., Adeyemi, I., Kumpf, B., Proud, E., Dror, I., Barrett, C. B., ... & Leeuwis, C. (2024). Innovation portfolio management for the public non-profit research and development sector: What can we learn from the private sector? *Innovation and Development*, *15*(3), 689-707.

Van Hoang, N. (2023). Human expertise and machine learning in collaborative intelligence frameworks for robust cybersecurity solutions. *Journal of Applied Cybersecurity Analytics, Intelligence, and Decision-Making Systems*, *13*(12), 1-12.

Wairagade, A., & Ranjan, S. (2025). Machine learning driven trust and reliability in human-AI collaboration for cybersecurity. In *2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS)* (pp. 1-8). IEEE.

West, G., & Hale, R. (2025). Building human-AI collaboration models in cybersecurity operations. *Proceedings of AI Applications held in Instabul 24 October, 2025.*

Yaich, R., Balondrade, A., Sicard, A., Fouquiau, C., Giraud, G., Amokrane-Ferka, K., & Arbaretier, E. (2025). Symbiotic human–AI collaboration for augmented cybersecurity operations. In *Proceedings of the AAAI Symposium Series* (Vol. 6, No. 1, pp. 350-358).

Declaration

Consent for publication

Not Applicable

Availability of data and materials

The publisher has the right to make the data public

Conflict of Interest

The authors declared no conflict of interest

Ethical Considerations

Not applicable

Competing interest

The authors report no conflict or competing interest

Funding

The author declared no source of funding

Authors' Contribution

APPENDIX I

DATA EXTRACTION TOOL

Human-AI Collaboration in Cybersecurity Decision-making: A Systematic Review of Literature

S/N	Research titles and authors	Aims	Methodology	Findings
1	Optimizing AI and human expertise integration in cybersecurity: Enhancing operational efficiency and collaborative decision-making Saadallah *et al*. (2025)	The study explored how artificial intelligence, automation, and human expertise can optimize cybersecurity operations through complementary roles and continuous feedback mechanisms	Qualitative approach was adopted.	- The symbiotic integration framework, which promotes continuous interaction and ethical oversight between AI systems and human operators, and the symbiotic maturity integration model, which maps the progressive stages of AI-human integration from initial to optimized. - The study showed that AI-human collaboration has the potential of guiding organizations in developing resilient, adaptive, and ethically responsible cybersecurity strategies as their roadmap advances on their maturity journey.
2	Building human-AI collaboration models in cybersecurity operations. West and Hale (2025)	The study explores the design and implementation of human-AI collaboration models to enhance cybersecurity efficiency, adaptability, and decision accuracy.	This study adopts a conceptual and analytical research design supported by secondary data and simulation-based validation.	- It examines frameworks where human expertise and AI-driven analytics coexist to improve threat detection, incident response, and risk management. - Through a study of hybrid intelligence paradigms, decision-support systems, and real-world cybersecurity applications, the study proposes a collaborative architecture that optimizes task allocation between humans and AI. - The paper highlights key challenges including trust, explainability, data privacy, and dynamic learning, and suggests strategies for achieving seamless cooperation between human operators and AI systems.

3	Human expertise and machine learning in collaborative intelligence frameworks for robust cybersecurity solutions. Van Hoang (2023)	The study explored the integration of these complementary strengths within a collaborative intelligence framework, emphasizing their role in building resilient cybersecurity systems.	Qualitative research approach.	- Human input refines ML models, handles ambiguous situations, and ensures ethical oversight, while ML automates repetitive tasks, detects real-time threats, and analyzes vast datasets. - Proposed solutions include employing explainable AI (XAI), iterative feedback loops, and prioritization algorithms to optimize human-machine collaboration. Finally, the paper explores future directions, including preparation for emerging threats like quantum computing and IoT vulnerabilities.
4	Symbiotic human–AI collaboration for augmented cybersecurity operations. Yaich *et al*. (2025)	The study proposes a human-AI collaboration framework to augment Security Operations Centres (SOCs) effectiveness through cognitive profiling and agentic coordination.	The study introduces a multi-agent architecture grounded in the Belief–Desire–Intention (BDI) model and structured by an extended VOWEL+U framework that embeds human oversight into agentic ecosystems.	- The study extended the VOWEL framework to include human agency (VOWEL+U), creating a foundation for cohesive agent–human collaboration. The framework enables immediate augmentation of high-friction workflows and longer-term transformation toward adaptive, trustable, and symbiotic cyber-defence teams. - In contrast, our work profiles SOC functions across three cognitive dimensions (thinking mode, attention demand, coordination context), prescribes four finely graded agent roles (Assistant, Auto-Pilot, Companion, Operator) grounded in BDI transparency to enable dynamic mode selection and shared situational awareness across the entire SOC lifecycle.
5	Human-AI collaboration: Enhancing	The study explores the impact, benefits, and challenges of	The study adopted a qualitative research approach.	- Human-AI collaboration leverages the strengths of both human intuition and machine

	decision-making in critical sectors. Etuk and Omankwu (2025)	integrating AI with human decision-making across critical sectors.		intelligence to enhance accuracy, efficiency, and reliability in decision-making. - AI systems provide data-driven insights, predictive analytics, and automation, while human expertise ensures ethical considerations, contextual understanding, and adaptability. This synergy improves risk assessment, crisis management, and strategic planning, ultimately leading to more informed and effective decisions. - However, challenges such as trust, transparency, and bias in AI models must be addressed to maximize the benefits of human-AI collaboration.
6	Human-AI collaboration and cyber security training: Learning analytics opportunities and challenges. Maennel and Maennel (2024)	The propose a holistic human-AI interaction model within the learning analytics (LA) and cyber security exercises (CSX) context.	The study adopted qualitative research approach.	- The model brings together elements and processes of human-AI interactions, as well as cyber ranges, cyber security, and LA tools, and a wider lens of multimodal learning analytics, exercise life-cycle, and overall pedagogical approach.
7	Enhancing AI-human collaborative decision-making in Industry 4.0 management practices Alam and Khan (2024)	The study focused on using AI-human collaborative decision-making in Industry 4.0 management practices.	The proposed approach provides a novel framework, stepping towards the enhancement of AI-human communication with real-time feedback, iterative refinements, and user-centric interface designs.	- The study showed that the indicated case studies and application scenarios will show the applicability and effectiveness of the framework in different contexts of industry, and therefore, provide concrete examples of how the benefits of the framework might be in practice. - Simulation results show that the proposed mechanism has been significantly adopted, demonstrating a 10-20% increase in efficiency, user satisfaction, and feedback

				responsiveness compared to the existing mechanisms such as EHIDM and HCADMR. - The results underscore the potential of the proposed framework to significantly enhance interaction dynamics between AI systems and human users.
8	Explainable AI in Cybersecurity: A Comprehensive Framework for enhancing transparency, trust, and Human-AI Collaboration Desai *et al*. (2024)	This comprehensive framework explores the application of XAI techniques like SHAP, LIME, and EBM to address the challenges of explainability and trustworthiness in AI-driven cybersecurity solutions.	The study incorporates Explainable AI (XAI) methodologies into cybersecurity structures.	- The study offers understandable explanations about how AI systems reach their conclusions. This empowers cybersecurity professionals to understand, validate, and effectively respond to cyber threats. The evaluation of XAI techniques against traditional cybersecurity models reveals superior performance, particularly in intrusion detection systems (IDS), phishing detection, incident response, vulnerability assessment, threat intelligence, anomaly detection, malware detection, SOAR, and user behavior analytics (UBA).
9	Evaluating human-machine interaction paradigms for effective human-artificial intelligence collaboration in cybersecurity Malatji (2024)	The study examines the efficacy of various Human-Machine Interaction (HMI) paradigms in enhancing cybersecurity practices through human-artificial intelligence (AI) collaboration.	Six Human-Machine Interaction (HMI) paradigms are analyzed: Humans in the Loop (HITL), Humans on the Loop (HOTL), Humans out of the Loop (HOOTL), Humans alongside the Loop (HATL), Humans-in-command (HIC), and Coactive Systems.	- HITL is about active direct human intervention while HOOTL emphasises autonomous AI operations. HOTL balances AI autonomy with human oversight. In HATL, AI and humans work simultaneously on different tasks, whereas in Coactive Systems, humans and AI collaborate equally and interdependently. Lastly, in HIC, humans are the final decision-makers and can override AI decisions. The strengths and weaknesses of the six HMI paradigms are determined by evaluating their key components against high-level cybersecurity practices, leveraging the advanced capabilities of ChatGPT-40.

				- The findings underscore the need for a hybrid approach that flexibly integrates multiple paradigms to optimize performance. Recommendations for practical implementation are provided, along with an outline of areas for future research, including real-world testing and the exploration of emerging AI advancements.
10	Augmented intelligence framework for human–artificial intelligence teaming in cybersecurity. Malatji (2025)	The study introduces the cybersecurity Augmented Intelligence Framework (*c*AIF), a conceptual framework designed to optimise human-AI teaming (HAIT) in cybersecurity.	Qualitative research approach.	- The analysed data suggests that strategically leveraging the strength of each paradigm allows for a hybrid intelligent framework comprising five core components: the Decision-Making Matrix, Paradigm Allocation Engine, Task-Specific Modules, Feedback and Learning System, and Interoperability Framework. The *c*AIF shows promise in enhancing human-AI collaboration, integrating human insights with AI capabilities to improve resilience and adaptability against evolving cyber threats.
11	From black box to trustworthy AI: A secure framework for explainable cybersecurity decision-making. Safavi *et al*. (2025)	The study seeks to address this particular constraint by introducing an innovative secure framework aimed at elucidating AI processes within the cybersecurity context.	The proposed framework amalgamates various explainability techniques, including SHAP, LIME, and Grad-CAM, to elucidate the rationale underlying AI decision-making.	- Results indicate comprehensive security layers that encompass data safeguarding, adversarial mitigation, and model provenance to uphold the integrity and robustness of AI systems. Moreover, trust-enhancing mechanisms, which include thorough logging, human-in-the-loop supervision, and model certification, are incorporated to cultivate confidence in AI-driven cybersecurity determinations. This framework aspires to augment transparency and reliability in essential security applications such as malware identification, phishing remediation, and AI-

				facilitated Security Operations Centers (SOCs), thereby paving the path for a more credible and efficacious application of AI within the cybersecurity paradigm.
12	Collaborative human-AI decision-making systems. Dolgikh and Mulesa (2021)	The study investigated collaborative human-AI decision-making systems in cyber	The study adopted multi-channel decision-making system.	- The proposed parallel multi-channel architecture combining human and machine expertise into a single synergetic system offers a number of essential advantages over conventional “single-chain” decision-making models, including: a significant improvement in overall accuracy of decisions. For two-stage systems, it does not introduce additional delays in the decision process due to high operational capacity of the machine intelligence channel, whereas for single-stage ones, offers significant improvement in performance; flexibility: the system is highly adaptable and transferrable to different areas / domains of application; it allows optimal use of limited expert resources only in the situations that require expert attention; is fully compatible with distributed, high-performance performance models of service delivery; combines strengths and advantages of the human and machine intelligences for an optimal outcome; and allows to retain complete human control over critical decisions.
13	Machine learning driven trust and reliability in human-AI collaboration for cybersecurity Wairagade and Ranjan (2025)	The study explores the challenges of human-AI collaborations in cybersecurity by	The analysis employed several Machine-Learning models, including Logistic Regression and Random-Forest; and Natural Language Processing	- The encouraging findings indicate that a medium trust level of 36 % can result in a 70 % success rate of Human-AI collaboration. Furthermore, feature importance analysis highlights that reliability scores, explainable AI (XAI), continuous model validation, and robust

		focusing on trust and reliability.	(NLP) techniques, along with Latent Dirichl*et al*location (LDA).	feedback loops will build a more effective and trustworthy Human-AI collaboration in future cybersecurity.
14	Human factors in AI-driven cybersecurity: Cognitive biases and trust issues. Hagen *et al*. (2025)	This study investigates how cognitive biases, such as automation bias (47%) and confirmation bias (37%), influence security analysts' trust in AI-driven tools, drawing on Kahneman's dual-process theory.	Through qualitative interviews with 19 cybersecurity professionals and a comparative analysis of AI solutions from Microsoft, CrowdStrike, Darktrace, and IBM, we identify key barriers to adoption, including explainability gaps and high false positive rates.	- Findings reveal that 65% of analysts express skepticism toward AI alerts, favoring hybrid human–AI models (79%) over full automation. We propose strategies like Explainable AI (XAI), bias-awareness training, and adaptive trust calibration to mitigate biases and foster trust. These insights highlight the need for user-centric AI designs that balance technical performance with human cognitive realities in cybersecurity operations.
15	Human-machine cooperative AI decision-making with heterogeneous data Blasch *et al*. (2023)	The study explores the elements associated with the opportunities and challenges emerging from designing, testing, and evaluating such future systems.	The study highlights the MAST (multi-attribute scorecard table), and more specifically the MAST criteria —analysis of alternatives‖ by measuring the risk associated with an evidential DL-based decision.	- The study demonstrated that the idea of risk includes the probability of a decision as well as the severity of the choice, from which there is also a need for an uncertainty bound on the decision choice which the paper postulates a risk bound. Notional analysis for a cyber networked system is presented to guide to interactive process for test and evaluation to support the certification of AI systems as to the decision risk for a human-machine system that includes analysis from both the DL method and a user.
16	Human-in-the-loop intelligence: Advancing AI-centric cybersecurity for the future. Karunamurthy *et al*. (2023)	The primary focus of the research revolves around the deep integration of artificial intelligence into	Qualitative research approach, using the Human-in-the-Loop Intelligence Cybersecurity Model introduces an	- Results showed that the conceptualization of this model is deeply rooted in a holistic understanding derived from it, which reflects the incorporation of the latest algorithms and techniques. By embracing the most recent advancements, this contribution

		crucial areas of cybersecurity, which includes activities such as authenticating user access, enhancing awareness of network situations, monitoring for potentially harmful behaviour, and identifying irregular traffic patterns.	innovative conceptual model.	offers a forward-thinking perspective to the ongoing discourse in AI-centric cybersecurity, thereby positioning itself at the forefront of this dynamically evolving field. - The symbiotic interplay between AI experts, cybersecurity specialists, ethicists, policymakers, and communication professionals is not just a strategic choice but a necessity. The convergence of diverse expertise is crucial not only for technical problem-solving but also for developing ethical frameworks, effective policies, and transparent communication strategies.
17	AI-driven cybersecurity: Opportunities, challenges, and the future of human-AI collaboration. Gyles and Boonthum-Denecke (2025)	The study investigates the strengths and limitations of AIdriven cybersecurity by comparing AI-based security tools with traditional methods, identifying key advantages and vulnerabilities, and exploring ethical considerations.	Experimental research approach.	- AI significantly enhances cybersecurity but must be continuously refined to remain effective. - Human expertise is irreplaceable, as AI lacks contextual understanding and decision-making skills. - AI security systems are vulnerable to adversarial attacks, requiring stronger safeguards and transparency. - Cost remains a challenge, but AI's scalability makes it a promising long-term investment. - Overall, the results suggest that AI is best used as an augmentative tool rather than a standalone solution. As technology evolves, integrating AI with human intelligence will be the most effective approach to strengthening cybersecurity.