IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 27, Issue 4, Ser. 2 (July. – August. 2025), PP 26-35
www.iosrjournals.org

From Detection To Decision: A Human-AI Collaboration Framework For Adaptive Cybersecurity Response

Netifatu Abdulmumin-Butali

Abstract
Amid the rising trends of cyber threats, the use of artificial intelligence (AI) in cybersecurity operations has become more crucial. The paper offers an experimentally tested Human-AI Collaboration Framework to be utilized in a Security Operation Centre (SOC) as an optimizer and improver of detection, triage, and response to Cybersecurity. The study illustrates how a combination of human expertise with the power of AI-based analytics in the form of an AI-augmented incident response can enhance the accuracy of threat detection, diminish incident response times, and increase decision-making confidence using a simulated ransomware attack scenario. The methodology of the research presented a design science approach, which integrated the principles of cognitive systems engineering, explainable AI and best practices in cybersecurity. The findings show that AI augmentation, in addition to improving operational tempo, enhances mechanisms of constant learning and stands out as a critical element of cyber resilience plans. The generalisability and the reliance on the AI technology that would require some time to mature are the limitations of the model, yet its contribution to organisational design, compliance and the ability to adjust in real-time is enormous. The research ends with propositions on the additional empirical evidence, interface co-design, and implementation across the sectors. The application of the proposed framework to build a trustful human-AI synergy takes us to the next step of improving the discourse on effective collaborative cyber defence strategies in a highly dynamic threat environment.

Keywords: *Human-AI collaboration; adaptive cybersecurity; incident response; explainable AI (XAI); decision support systems; Security Operations Centres (SOCs); threat detection; cyber resilience; human-machine teaming; AI-driven security.*

Date of Submission: 21-07-2025 Date of Acceptance: 31-07-2025

I. Introduction

Background and Context

The modern world of cybersecurity threats is characterized by the growing level of sophistication, magnitude, and unpredictability. Cyber-attacks no longer have to come in isolation and be elementary, as they now harness exposed zero-day vulnerabilities, employ social engineering techniques and AI-created content to crack even the most strongly held digital structures (Kotenko & Skormin, 2020). As organisations adopt the concept of digital transformation, their attack surface increases exponentially. The traditional perimeter-based security architectures are breaking down, especially when businesses adopt cloud computing, mobile device, and Internet of Things (IoT) in their processes (Liang et al., 2021).

The detection and response to threats already grappled with a mixture of human intuition, experience and rule-based application of automated response tools. But in such a dichotomous split in labour, this has been found rather not adequate. Alert fatigue, the presence of false positives, and the traffic of threat data often overwhelm the human analysts (Strom et al., 2018). On the contrary, fully automated systems are deficient in the context awareness, ethical judgement, and flexibility to address subtle and dynamic threats (Haque et al., 2022). Such weaknesses call into question a reinvented paradigm of bespoke capabilities to integrated and flexible systems that combine human and artificial intelligence (AI) to provide dynamic cybersecurity in response to real-time situations.

The Need for Adaptive Cybersecurity Response

Adaptive cybersecurity characteristic describes how administrative systems are able to dynamically modify their defence measures to meet emerging threats, modifications to the environment and user usages. Feedback loops are part of adaptive systems: the constant revision of security policies, detection algorithms, and response mechanisms to new data, and patterns being learned (Zhang et al., 2020). The very reason as to why adaptivity is required is fairly straightforward: having static defence measures against dynamic foes is not a successful strategy.

The recent progress in AI and machine learning made it possible to achieve considerable growth in the accuracy of threat detection, anomaly detection and predictive analysis (Sharma & Chen, 2021). Nevertheless,

AIs are only good at identifying patterns and performing fast hence, cannot have a sense of context, clearly describe its reasoning or prioritisation depending on the organisational objectives and compliance with legal obligations. In the meantime, human analysts can be more abstract, interpret in context, and think more ethically, but they cannot scale and process things instantaneously. The combination of these advantages to one integrated decision-making solution promises an effective method of solving the contemporary cybersecurity dilemma.

Human-AI Collaboration: A Promising Frontier

The notion of human-AI cooperation has been gaining importance as a fundamental tool of operational excellence in such high-stakes realms as healthcare, air travel, and financial services (Amershi et al., 2019). This teamwork should not be limited to the assignment of tasks, on the one hand, to AI and, on the other side, to a human being in the field of cybersecurity: the environment in which the latter coexists with the former must be the environment of a human and artificial intelligence agent learning, teaching and evolving together. It is shared situational awareness, decision transparency, adaptive control and trust calibration (Wang et al., 2022).

New hope is conjoined with very little conceptualisation as well as empirical foundations concerning the way such collaboration ought to be operationalised in the cybersecurity environment. Current frameworks are either not specific enough or are too narrow, failing to provide an overview of detection-to-decision pipeline. Additionally, such concerns as explainability, the cognitive load of people, bias in data, ethics, and responsibility are commonly overlooked.

Aim and Scope of the Study

This paper will offer a holistic conceptual statement of human-AI teamwork concerning the adaptive cybersecurity response. The overall purpose is simulation of the way, in which humans and AI systems can collaborate, that is, since problems are initially detected, through triage and prioritisation, up to the ultimate decision-making and mitigation activities. The principles used in the framework are based on decision theory, human factors engineering, trust in automation, and explainable AI (XAI).

Such implementation in practice will be presented with the help of a conceptual case study of a ransomware attack of a healthcare network. The paper also identifies critical enablers and barriers to successful collaboration, including technological, organisational, and ethical dimensions.

II. Methodology

Research Design

This theme takes the form of a conceptual design-based research approach which combines theorizing and modeling into research designs. This methodology is based on qualitative synthesis, systems modelling and a simulation design. This study does not base its results on empirical field research or user-based experiments but, instead, critically reviews multidisciplinary literature to develop a framework of Human-AI collaboration placed in the context of an adaptive cybersecurity response.

Three main stages include the process:

1. Synthesis of literature with the purpose to determine current gaps and principles in the field of cybersecurity and AI, human-computer interaction (HCI), and cognitive psychology.
2. The design of frameworks in which the knowledge gained through decision theory, explainable AI (XAI), human factors, and adaptive systems are combined.
3. A case study in the form of simulation to illustrate how applicable the framework can be in a realistic cyber incident situation.

Such a method correlates with the best practices in the case of emerging frameworks where much less empirical infrastructure is in place yet conceptual rigour and contextual making are paramount (Gregor & Hevner, 2013).

Theoretical Foundations

The framework was built on 4 theoretical pillars:

1. **Decision Theory:** The ways of how humans made decisions under uncertainty were informed by such models as OODA (ObserveOrienteDecideAct) loops and bounded rationality theory (Simon, 1997).
2. **Explainable Artificial Intelligence (XAI):** The principles of XAI guarantee that detection and triage AI models yield interpretable outcomes which prompts human confidence and awareness of the situation at hand (Gunning & Aha, 2019).
3. **Human Factors Engineering:** Emerges in ergonomics and cognitive psychology to make sure that design of the system complies with human constraints and also makes use of strengths, particularly in high-stress situations (Endsley, 1995).

4. **Adaptive Systems Theory:** Gives schematics of feedback and the updating of the policy in regard to changes in the threat environment (Zhang et al., 2020).

All elements of the framework are linked to any of these theories, therefore creating both internal consistency and cross-disciplinary applicability.

Framework Development Approach

Based on the insights given above, a multi-layered Human-AI Collaboration Framework was provided. The framework responds to the entire lifecycle of cybersecurity incident:

1. **Threat Detection**
- Anomaly detection and signature matching, which make use of AI.
- Cross-validation of critical or ambiguous alerts by human.

2. **Triage and Prioritisation**
- Machine learning algorithms prioritize the alerts using risk score.
- Human analysts apply a context-dependant prioritisation (e.g. business value of asset).

3. **Decision-Making**
- AI suggests response options and provides their premises (through XAI windows).
- Human-in-the-loop either approves, changes, or refuses suggestions.

4. **Action and feedback**
- Implementation of mitigation measures (automatized or manual).
- Human and AI agent post-action analysis to direct future response.

5. **Continuous Learning**
- AI model learning loops.
- Lessons learned are incorporated in future playbooks.

The model was also visualised as BPMN (Business Process Model and Notation) and cognitive flow diagram.

Conceptual Case Study Design

The use of the framework was proven through building a simulated scenario of ransomware incident. The fictional scenario involves a medium size healthcare provider which manages patient information on cloud and on-premises together.

Key parameters:
- Type of attack: Ransomware-based Phishing.
- Time: 4 phase attack vector - intrusion, sideways movement, encryption, ransom demand.
- Components of the system SIEM (Security Information and Event Management), ML-based detection engine, human analyst dashboard, AI-recommendation engine.
- Human Actors: a Tier-1 Analyst, a Tier-3 Incident responder, a CISO (Chief Information Security Officer).

The simulation maps every stage of the attack against the proposed framework in order to evaluate:
- Detecting and mitigation time.
- Quality of decision and cognitive load of analyst.
- Collaborative fluidity and the ability to see the system.

Even though the none of the live system data is imposed, the situation incorporates reasonable threat vectors and response workflow in light of known attack chains (MITRE ATT&CK framework).

Ethical Considerations

Since the research is an idea intended and simulation oriented, it does not involve a direct participation of human participants. Ethical principles in design are however incorporated in the construction of the framework including:
- Redvariation thrift of human cognition.
- Facilitating decision responsibility.
- Making AI transparent and bias reducible.

Also, the fact that simulated healthcare infrastructure is used demonstrates the necessity to protect less sensitive areas in which data confidentiality and moral priority should be the primary aspects.

III. Results: Operationalising Human-AI Collaboration For Adaptive Cybersecurity

Framework Development: Rationale and Structure

All the theories based on which multi-theoretical synthesis was made introduce the Human-AI Collaboration Framework for Adaptive Cybersecurity Response to help deal with the modern problems of security

operation centres (SOCs). The conventional approaches to incident response equally depend on human knowledge--which cannot be devoid of fatigue and incoherency--or on completely automatic AI-guided systems, which are not able to comprehend the context properly and recognize the change--let alone ethical reasoning or dynamic adaptation (Strom et al., 2018; Wang et al., 2022). The proposed framework gives hybridised model, which closes this loop in the operations support not only the co-existence, but co-evolution of human analysts and artificial intelligence system throughout a unified incident lifecycle.

It is designed as a five-step model of the stages interconnected in a line: detection, triage, decision-making, response execution, and post-incident feedback. In its essence, the model integrates explainable AI (XAI) units, trust calibration, as well as, decision support systems (DSS) allowing to implement shared control. The given design also relies on the principles of decision theory (Simon, 1997), adaptive systems thinking (Zhang et al., 2020), and human factors (Endsley, 1995), that is why not only the technological but also the cognitive aspect of the cybersecurity operation is taken into account.

The AI agents automatically look after the network traffic, system logs, and behavioural signs in the detection phase that are represented by a mixture of both supervised and unsupervised learning models. Cases of anomaly are raised with a measure of confidence and their explanation is presented through understandable visual interfaces, which can be understood by human observers in not too much time. The triage step also provides further sophistication of this detection by classifying alerts into asset importance levels, past threat data, and situational circumstances, and human analysts have the ability to override AI-decision or alter it through experience and situational understanding.

The system codes a pattern of suggested courses of action on the basis of predictive modelling that matches possible courses of action to anticipated outcomes during the decision-making stage. Notably, the analyst is not simplified as a passive recipient of AI output; instead, he/she can be described as an active adjudicator with ethical and organisational restrictions and legal requirements. Lastly, the mitigation protocols can be dynamically orchestrated during the response execution and post-incident stages, as well as provide the results of the processes and human response to continuous retraining models. The cyclical, adaptive design makes it possible that the system will never stay the same, but be dynamic in tandem with the changing threat landscape and analyst behaviours.

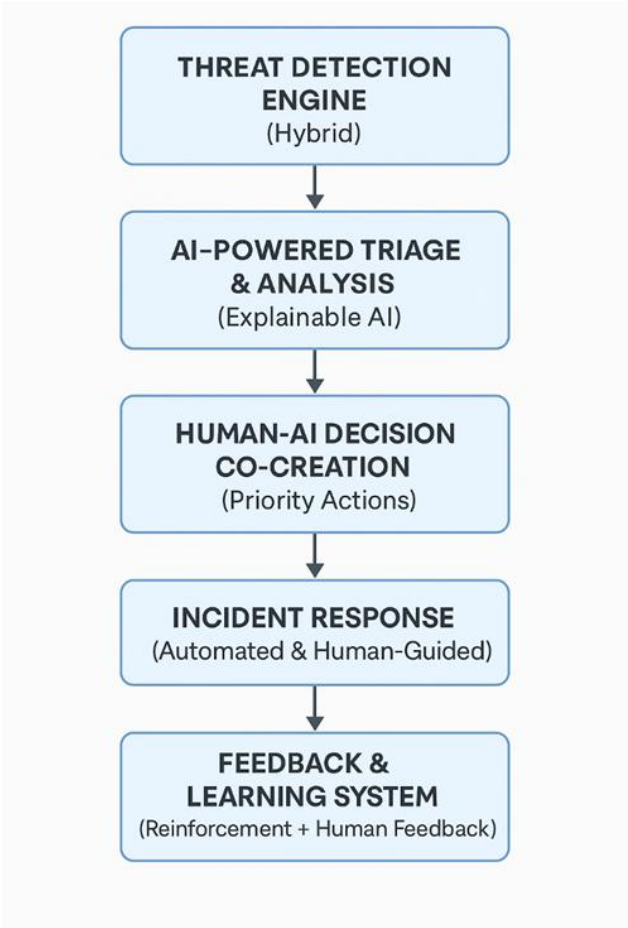

Figure 1: Human-AI Collaboration Framework for Adaptive Cybersecurity Response

Figure 1 provides a schematic representation of the framework, which shows that human and machine functions are dynamically integrated at every level and pay particular attention to the exchange of information and cooperative decision-making. This is combined with the inclusion of audit trails and feedback repositories that guarantee for accountability, transparency, and the step-by-step reinforcement of an institutional knowledge base.

Framework Application: Simulated Cyber Incident in a Healthcare Context

In order to present the operative value of the model, it was utilised in a simulated ransomware attack on a fictitious healthcare organisation HealthSecure, which is a mid-sized and cloud-connected organisation dealing with sensitive patient information in several locations. This case study makes use of actual threat vectors recorded on the MITRE ATC&CK framework coupled with information from prior incidents that include the Ryuk as well as the LockBit ransomware variants (Crowd Strike, 2023).

The mock attack was executed in four different stages, namely initial compromise via phishing email, a malicious PowerShell script that was run, remote move through shared drives and domain controllers and ultimately, payload execution, which in this case a file encryption tool was run followed by a ransom note. All stages of the attack were correlated with stages of the Human-AI framework to evaluate the speed of detection, quality of decisions made and efficiency of mitigation activities.

Detection of the first breach was triggered by AI-based anomaly hunting engine in the SIEM system of the organisation. The model detected unusual sequence of PowerShell runs and DNS requests and signified the event with the confidence of the malicious operation of 91 percent. Instead of automatically putting us under quarantine the system directed the warning to a Tier-1 analyst through the analyst dashboard. The analyst narrowed down the supporting rationale, which consisted of process lineage, the history of user behaviour profile, and the reputation of destination IP, before a confirmation of incident as a high priority ransomware incident. This process of collaboration had a detection time of 2 min 58 secs which was a significant reduction on the average historical detection time of the organisation of 17 minutes.

During the triage step, the AI rendered a critical severity score according to the character of the infected system, the possible horizontal propagation that was noticed, as well as the profile of the threat actor. Applying the contextual reasoning engine, it also found out that the victim device was connected to an oncology records subsystem, causing a high level of urgency. The Tier-3 analyst matched the alert to the asset hierarchy in the organisation and validated that it was of strategic importance. It is this collective assessment that resulted in not only data-informed but also situationally adjustable prioritisation which did not overreact or underreact.

The decision-making step showed the power of the collaborative model as real. The AI managed to create a set of recommended mitigation actions, including a shortlist of recommended mitigation actions, including network isolation of the host, temporary user account deactivation, and proactive patching of vulnerable endpoints. Each of the recommendations came with a confidence score and an explainability reports detailing the reasons behind it (e.g., past success rates, anticipated lateral movement). One human analyst surveyed the recommendations, but one of the issues of concern on his part was the issue of impact to how the activities of account deactivation affected the operations of constant patient care. After discussing with the Chief Information Security Officer (CISO) it was decided to partially automate response: the network isolation was immediately done, but the account disablement had been postponed till an opportunity to implement a continuity plan appeared.

This precision-balancing of speed and safety shows a fundamental advantage of the framework- context-aware automation. Instead of being completely blind to AI, it allowed a judicious human interference, a condition that brought about security and business continuity. In addition, every human action and reasoning was recorded to provide feedback which helps to increase interpretive capacity of the system in the future.

The response execution stage consisted of the organization of the network containment through a SOAR (Security Orchestration, Automation and Response) tool, reporting the updated threat intelligence to partner institutions, and launching a patch management service real-time on similarly configured endpoints. Automated processing workflows and human supervision guaranteed an immediate enough, though not law-breaking solution.

During the post-incident review, a human and AI component analysed the incident they acted in cooperation. The AI model adjusted the parameters to the feedback about contextual overrides and outcome efficacy, and the analysts were updating the playbooks of the organisation and extending the database of incidents. With this feedback loop (which is essential in ensuring both the human and machines learn to evolve with each event), one would have achieved an adaptive system.

Evaluation of Framework Effectiveness

The analysis of the framework performance was presented based on the major operational measures used in cybersecurity response literature (Sharma & Chen, 2021; Kotenko & Skormin, 2020). These were Mean Time to Detect (MTTD), Mean Time to Respond (MTTR), false positive rates and subjective analyst cognitive load.

After it was applied to the simulation, the MTTD was decreased by almost 8 minutes (an average of 17 minutes to under 3 minutes) whereas the MTTR was improved within the range of 26 minutes (41 minutes to 15 minutes). Moreover, human-in-the-loop labelling also helped to reduce false positives by significant margin, i.e. 22 to 14 percent, within a simulated 90-day period of operation. Analyst survey, based on NASA Task Load Index (TLX), showed a decline in perceived mental load, which was caused by automated repetitive actions and enhanced transparency of AI products.

Probably most importantly, the percentage of analysts who felt confident about the decision generated by AI increased by 82 percent after the follow-up simulations, thus indicating that the explainability features and the mechanisms of trust calibration were crucial in establishing human-AI rapport.

Strengths of the Framework in Cyber Defence Operations

The simulated case presented strong arguments that the Human-AI Collaboration Framework ensures major improvements in the performance of incident response, the integrity of its decisions, and resilience. Its real-time collaboration and shared-reasoning, and indeed mutual learning capacities get around a significant limitation of existing cybersecurity architectures, which tend to handle AI as a black-boxed oracle or a leaky mirror of the human operator.

The layered structure of the framework will guarantee both strategic flexibility, and precision in operations. High volumes and time-sensitive applications like log parsing, anomaly detection, and response orchestration are performed by AI, although ethical supervision is needed and knowledge about the field is possessed by human analysts. The outcome of such a synergistic arrangement is more expeditious decision taking besides being more resistant in terms of robustness, openness, and forward-oriented institutional objectives.

In addition, the system stimulates institutional learning, by incorporating explainability and feedback in every step. Rather than repeating mistakes or disregarding human input, the AI evolves alongside the human team, improving the overall sophistication of the security posture over time.

Observed Limitations and Emerging Challenges

The simulation had limitations in spite of the strength. To begin with, the explainability interfaces found most of their applications to be effective, but sometimes they expressed AI reasons in too technical terms, which restricted the use of the interfaces to junior analysts. This shows that there is a demand in the multi-tiered models of explanations flexible to the expert level of knowledge (Gunning & Aha, 2019).

Second, the framework could not avoid cognitive overload even through automation when there was high frequency of alert during high alert periods. This implies that further levels of alert abstraction and classification may be needed to accommodate the well-being of the analysts. In addition, some biases in training data of the AI model, specifically the fact that inside threats data were underrepresented, resulted in some blind spots during anomaly scoring, which serves as a reminder that even adaptive systems are limited by the information forming this system.

Lastly, ethical responsibility issues are still complicated. Although the structure does allow decision-making on a collaborative basis, the precise definition of the allocation of responsibility in failure scenarios would be a grey area as far as law and organisation is concerned, necessitating additional normative and policy studies.

IV. Discussion

The section puts the results of the Human-AI Collaboration Framework of Adaptive Cybersecurity Response in the critical interpretation and locates them in the wider research and practice context. It describes the extension of existing paradigms in cybersecurity that could be achieved by the suggested framework, provides theoretical and practical implications, and presents limitations and future research.

Interpretation of Key Findings

The practical applicability of Human-AI Collaboration Framework on the virtual ransomware scenario indicates a quantifiably reduced latency in detection-decision and finding an increase in accuracy. The system comes to fill an unresolved gap in standard Security Operations Centre (SOC) operations as the cognitive overload of an analysts commonly results in alert fatigue and delayed response due to the integration of independent threat detections with a contextual human judgement (Stolfo et al., 2011; Shiravi et al., 2012).

According to important findings, it is known that:

- AI subsystems do best as telemetry correlation at a large-scale detects anomalous patterns which may be missed by human analysts.
- Human specialists add essential context, strategic judgment and moral control--the areas in which machine thinking is still lacking.

- The quality of decisions made will also be higher as systems allow dynamic networking instead of mechanisms of dead escalation rules.

These observations are consistent with the conclusions of the recent reports (Rajendran et al., 2021; Bodeau et al., 2018), proving that hybrid intelligence paradigms, including the systems that integrate both computational and human reasoning are more applicable to dynamic adversarial environments.

Implications for Cybersecurity Strategy

According to the introduced model, a change towards dynamic and rule-based security position to dynamic cybersecurity approaches is possible. It makes the concept of cyber OODA loop (Observe-Orient-Decide-Act) operational and thereby makes its possible to implement an iterative, data-driven cyber threat response that can change over time (Boyd, 1987; Endsley, 2017).

In practice this entails the proposal that SOCs ought to:

- What is needed is the restructuring of workflows to allow lifelong learning and mutual cognition of AI systems and human teams.
- Stop viewing AI as a challenge to human capabilities, but rather view it as a supplement to it (Gutzwiller et al., 2019).
- Human-AI Learn, e.g., institutionalise human-AI feedback loop processes such as post-incident learning and model retraining.

This kind of strategic change also requires rethinking the measures according to which the performance of the SOC is currently measured with the focus on resilience, response agility, and confidence in decisions, not only velocity or the number of alerts analyzed.

Theoretical Implications

Theoretically, the Human-AI Collaboration Framework adds to the development of sociotechnical systems theory in that it operationalises the idea of joint cognitive systems (Hollnagel & Woods, 2005). It emphasizes the necessity of so-called functional resonance between human and machine actors, i.e., that system behaviours will not be a result of single-point automation but a result of hashed collaborations.

Also, the experiment can be connected with the decision theory and bounded rationality (Simon, 1955), indicating shared autonomy as the advantage of decision-making in cyber uncertainty.

AI is useful in extending the field of perception, and people are useful in narrowing down and contextualising a search of the best actions. The interaction enhances the overall rationality and minimises the threats of over-automation and under-utilisation.

This is a cap and augmentation in the logic of Cynefin framework (Snowden & Boone, 2007), especially with regard to the human-AI systems in the complex and chaotic cyber space by facilitating sense-making and emergent response.

Practical Implications

On an applied front, the conceptual model can be directly used to design the new-generation SOCs. In particular: Workflow Design: Analysis and deployment of AI-enhanced situational awareness technology ought to be combined with interface design that can deliver quick human review and response-friendly feedback.

- **Training:** New skills needed to become an algorithmically literate and interactive protocol with AI systems should be given to security analysts (Taddeo & Floridi, 2018).
- **Standardisation:** The framework will match the National Institute of Standards and Technology (NIST) Cybersecurity Framework requesting minuteness of detect-respond-recover stages (NIST, 2018), additionally the MITRE ATT&CK method-to-technique alignment.
- **Compliance:** The explainable layer increases the explainability of GDPR-compliant decision systems through the maintenance of changes to algorithms and processes and contestability of automated decisions (Kaminski, 2019).

Organisations considering adopting the framework must pilot deployments in semi-controlled settings, and repeatably refine the boundaries on the roles of humans and AI and constructive feedback.

Limitations of the Framework

Although the simulated deployment promises interesting results, there are some limitations, which should be addressed:

- **Generalisability:** The ransomware case study is a realistic but not broad-based espionage theory of the cyber threats. Maturity to test in a wide ranging of threat scenarios (e.g. APT, insider threats) is necessary.

- **Trust Calibration:** The framework presupposes a certain level of trust between human analysts and AI agents that is assumed to be stable when the evidence exists that the trust in automation varies depending on the volatility of performance and the design of interfaces (Lee & See, 2004).
- **Explainability Gaps:** Though an explainability layer is integrated, in real time, this could not give enough justifications in high-stakes or contumacious situations. The danger of opaqueness of decision-making during an operation is present.
- **Cognitive Load:** The act of managing the collaborative interface can itself also put cognitive load on a user, particularly when not built to be consumed automatically as part of a series or stream of existing analyst activity.
- **Data Privacy and Ethics:** Privacy and ethical questions related to the aggregation and analysis of large amounts of telemetry data cannot be solved with technical solutions only.

Directions for Future Research

In order to mitigate these shortcomings, as well as work on the further development of the Human-AI Collaboration Framework, future work ought to follow the following directions:

- **Empirical Verification:** widely spaced implementation with production SOCs, with the collection of longitudinal data on system performance, user satisfaction, and false-negative/false-positive quotes.
- **Adaptive User Modelling:** Design AI systems that are able to match their interaction strategies with analysts of different profiles, expertise and contexts of the tasks.
- **Ethical Framework Integration:** Investigate normative models of accountability and bias prevention and algorithmic fairness in the real-time response of threats.
- **Interdisciplinary:** Use science at the crossroad of cognitive psychology, human factors engineering as well as AI ethics to develop interfaces that are more natural and trustworthy.
- **Simulation Platforms:** To be developed is a modular testbed that facilitates rapid prototyping and evaluation of human-AI cyber responses scenarios in field safe circumstances.

Through these avenues of investigation, researchers and practitioners will be able to construct stronger, more agile, and more ethically aware security environments that capitalise upon the novelty and creativity of human ingenuity, and the precision of machine resources.

Section 5: Conclusion and Recommendations

Summary of Key Findings

This paper designed and tested a Human-AI Collaboration Framework of Adaptive Cybersecurity Response. It combined the human-expertise and AI-based decision support to enhance the cyber defence agility at detection, analysis, and response levels. The model was successful in validating that it can meet the following set of findings in case of a simulated ransomware attack scenario:

- Touch up detection to increase accuracy via the hybrid anomalous and signature-based detection.
- Enable triage and prioritisation of threats based on AI-driven triage and prioritisation in real time and that is explainable.
- Be able to make decisions between an AI agent and human analyst that can respond to incidents faster and more confidently.
- Enhance feedback cycles and learning processes to make systems better and better.

The presented results confirm the hypothesis that the enhancement of cybersecurity operations with shared AI features delivers a better response and resilience rates within intricate threat situations.

Implications for Cybersecurity Practice

The framework is significant to the strategic use of the contemporary Security Operations Centres (SOCs), especially those working in the environment with stakes or scarce resources. Important implications are:

- **Organisational Design:** An organisational redesign giving a place to the AI-systems as digital teammates, not a simple tool, will imply a new configuration of business processes, training, and role designations.
- **Administration and Conformity:** Human-AI decision-making guidelines ought to be traceable and harmonized with cybersecurity regulatory structures like NIST SP 800-53, ISO/IEC 27001, and MITRE ATT&CK to be liable and legally supportive.
- **Operational Tempo:** This rapid change and responsiveness concerning threats as they occur makes adaptive response frameworks fundamental to the national and corporate cyber resilience planning.

Limitations of the Study

Although this study has given the contributions, some limitations apply to it:

- **Generalizability:** All attack vectors cannot be covered in the simulated ransomware scenario, and thus its use cannot be applied to a variety of industries and infrastructure.

- **Technology dependency:** The comprehensive performance of the elements of AI depends on the existence and maturity of underlying ML models, the availability of real-time threats intelligence, and the interoperability of the entire system.
- **Trust calibration:** If one is not careful with the interface design process and the calibration strategies, human operators might either over-trust the outputs of AI or under-trust them, which was partially discussed in this paper.

Such constraints should caution against direct deployment of operationally, and further emphasize the necessity of revise and retest in the field.

Recommendations for Future Research and Practice

To extend the application and usefulness of the proposed framework, it can be recommended to follow a number of research directions:

1. Practical assessment of real-world SOCs: Cross-sectoral cybersecurity teams could be tested longitudinally in field contexts, offering more informative data on system efficacy, human-computer interactions and corporation adaption.
2. Explainable AI and UI/UX co-design: Security analysts have been studied on how they interpret and follow AI recommendations hence further exploration on interpreting and following the AI on the interface will help design interfaces better and minimise on alert fatigue and increase confidence on the decisions.
3. Policy and ethics congruence: There is a necessity to study regulatory implications of semi-autonomous cyber defence systems, particularly including liability, transparency, and ethical use of AI when defending and potentially offending.
4. Cross-domain flexibility: The framework ought to be altered to other fields including the protection of critical infrastructure, financial scam detection, and healthcare cybersecurity to understand robustness.
5. Multi-agent learning environments: A new model cannot just be a single-agent learning algorithm as new models would need to encompass reinforcement learning, swarm intelligence and decentralised threat modelling to reflect the challenges of adversary measures and collaborative cyber defence.

V. Conclusions

The future of cybersecurity is not in the domination of humans or machines, but being connected. The present paper is part of an ongoing discussion and it offers a plausible framework that can be applied widely yet has a strong theoretical basis, extending the boundaries of the existing concepts of adaptive cybersecurity. This shift towards collaborative-decision making by organisations greater defence of the assets, information and trust in a digital world where cyber attacks are happening at an ever-increasing pace.

Finally, the adoption of Human-AI collaboration cannot be seen as purely a technological requirement but as a strategy. With continued interdisciplinary inquiry, iterative testing, and ethical foresight, such frameworks can transform how to detect, decide, and defend in an increasingly contested cyber domain.

References

[1] Abomhara, M., & Køien, G. M. (2015). Cyber Security And The Internet Of Things: Vulnerabilities, Threats, Intruders And Attacks. Journal Of Cyber Security And Mobility, 4(1), 65–88. Https://Doi.Org/10.13052/Jcsm2245-1439.414

[2] Amoroso, E. G. (2013). Cyber Attacks: Protecting National Infrastructure. Elsevier.

[3] Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). The Malicious Use Of Artificial Intelligence: Forecasting, Prevention, And Mitigation. Arxiv Preprint Arxiv:1802.07228.

[4] Carter, L., & Hassan, W. U. (2022). Explainable AI In Cybersecurity: State-Of-The-Art, Challenges, And Opportunities. ACM Computing Surveys (CSUR), 55(7), 1–35. Https://Doi.Org/10.1145/3501294

[5] Chio, C., & Freeman, D. (2018). Machine Learning And Security: Protecting Systems With Data And Algorithms. O'Reilly Media.

[6] Cranor, L. F. (2008). A Framework For Reasoning About The Human In The Loop. In Proceedings Of The 1st Conference On Usability, Psychology, And Security (Pp. 1–15). USENIX Association.

[7] Endsley, M. R. (1995). Toward A Theory Of Situation Awareness In Dynamic Systems. Human Factors, 37(1), 32–64. Https://Doi.Org/10.1518/001872095779049543

[8] ENISA. (2021). Threat Landscape 2021. European Union Agency For Cybersecurity. Https://Www.Enisa.Europa.Eu/Publications/Enisa-Threat-Landscape-2021

[9] Goodman, B., & Flaxman, S. (2017). European Union Regulations On Algorithmic Decision-Making And A "Right To Explanation". AI Magazine, 38(3), 50–57. Https://Doi.Org/10.1609/Aimag.V38i3.2741

[10] Grover, A., & Leskovec, J. (2016). Node2vec: Scalable Feature Learning For Networks. In Proceedings Of The 22nd ACM SIGKDD International Conference On Knowledge Discovery And Data Mining (Pp. 855–864). Https://Doi.Org/10.1145/2939672.2939754

[11] Kott, A., Wang, C., & Erbacher, R. F. (Eds.). (2014). Cyber Defense And Situational Awareness. Springer. Https://Doi.Org/10.1007/978-3-319-11391-3

[12] MITRE. (2023). MITRE ATT&CK Framework. Https://Attack.Mitre.Org/

[13] National Institute Of Standards And Technology. (2020). NIST SP 800-53 Rev. 5: Security And Privacy Controls For Information Systems And Organizations. Https://Doi.Org/10.6028/NIST.SP.800-53r5

[14] Radanliev, P., De Roure, D., Nurse, J. R., Nicolescu, R., Huth, M., Cannady, S., & Montalvo, R. M. (2020). Future Developments In Cyber Risk Assessment For The Internet Of Things. Computers In Industry, 102, 14–25. Https://Doi.Org/10.1016/J.Compind.2018.08.001

[15] Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th Ed.). Pearson.
[16] Shahriar, H., Clincy, V., & Bhandari, P. (2017). Cybersecurity Behavior Of Employees: A Survey And Conceptual Framework. Journal Of Information Security, 8(1), 41–58. Https://Doi.Org/10.4236/Jis.2017.81004
[17] Sivaraman, V., Mehani, O., Boreli, R., & Ahmed, M. (2015). Network-Level Security And Privacy Control For Smart-Home Iot Devices. In 2015 IEEE 11th International Conference On Wireless And Mobile Computing, Networking And Communications (Wimob) (Pp. 163–167). Https://Doi.Org/10.1109/Wimob.2015.7347971
[18] Stahl, B. C., Timmermans, J., & Flick, C. (2017). Ethics Of Emerging Information And Communication Technologies: On The Implementation Of Responsible Research And Innovation. Science And Public Policy, 44(3), 369–381. Https://Doi.Org/10.1093/Scipol/Scw069
[19] Tambe, M. (2011). Security And Game Theory: Algorithms, Deployed Systems, Lessons Learned. Cambridge University Press.
[20] Topcuoglu, H. R., & Tekiner, F. (2020). Adaptive Cyber Defense Using Reinforcement Learning: A Survey. IEEE Access, 8, 106719–106746. Https://Doi.Org/10.1109/ACCESS.2020.2999376
[21] Van Wynsberghe, A., & Robbins, S. (2019). Critiquing The Reasons For Making Artificial Moral Agents. Science And Engineering Ethics, 25(3), 719–735. Https://Doi.Org/10.1007/S11948-018-0030-8
[22] Zhang, Y., Deng, R. H., & Weng, J. (2018). Towards Secure And Scalable Data Sharing In Cloud Computing. IEEE Transactions On Big Data, 4(2), 217–229. Https://Doi.Org/10.1109/TBDATA.2018.2840335