Navigating Autonomy: Unveiling Security Experts' Perspectives on Augmented Intelligence in Cybersecurity

Neele Roch, Hannah Sievers, Lorin Schöni, and Verena Zimmermann, *ETH Zurich*

https://www.usenix.org/conference/soups2024/presentation/roch

This paper is included in the Proceedings of the
Twentieth Symposium on Usable Privacy and Security.

August 12–13, 2024 • Philadelphia, PA, USA

978-1-939133-42-7

Navigating Autonomy: Unveiling Security Experts' Perspectives on Augmented Intelligence in Cybersecurity

Neele Roch *ETH Zurich*
Hannah Sievers *ETH Zurich*
Lorin Schöni *ETH Zurich*
Verena Zimmermann *ETH Zurich*

Abstract

The rapidly evolving cybersecurity threat landscape and shortage of skilled professionals are amplifying the need for technical support. AI tools offer great opportunities to support security experts by augmenting their intelligence and allowing them to focus on their unique human skills and expertise. For the successful design of AI tools and expert-AI interfaces, however, it is essential to understand the specialised security-critical context and the experts' requirements. To this end, 27 in-depth interviews with security experts, mostly in high-level managerial roles, were conducted and analysed using a grounded theory approach. The interviews showed that experts assigned tasks to AI, humans, or the human-AI team according to the skills they attributed to them. However, deciding how autonomously an AI tool should be able to perform tasks is a challenge that requires experts to weigh up factors such as trust, type of task, benefits, and risks. The resulting decision framework enhances understanding of the interplay between trust in AI, especially influenced by its transparency, and different levels of autonomy. As these factors affect the adoption of AI and the success of expert-AI collaboration in cybersecurity, it is important to further investigate them in the context of experts' AI-related decision-making processes.

1 Introduction

The growing dependence on digital devices, services, and data for daily tasks by individuals, companies, and governments increases productivity but concurrently increases the vulnerability to cyberattacks. Cybercrime is growing exponentially, with organisations experiencing an average of 1248 attacks per week [10, 11]. This results in a fast-paced and demanding work environment for cybersecurity teams. Simultaneously, organisations face a significant shortage of cybersecurity experts (CSEs) to cope with increasing security-related demands. In 2022 the estimated cybersecurity workforce gap stood at 3.4 million jobs globally [42], and existing CSEs report high work stress levels, fear of burnout, and feeling set up for failure in a chronic state of work overload [4, 41]. While educating future CSEs is an essential task, it may take a long time and currently does not catch up with the increasing demands.

Hence, technical solutions, such as Artificial Intelligence (AI) are a promising approach to ensure the secure operability of IT systems, to compensate for the current lack of CSEs, and relieve experts. AI tools have the potential to support human CSEs by enhancing and strengthening the human's abilities to act, analyse, decide, see and hear [61]. Human-AI collaboration does not aim to replace humans, but "to achieve complex goals by combining human and AI, thereby reaching superior results to those each of them could have accomplished separately [...]." [25, p. 640]. This complementary collaboration has already been proven to be successful in other domains, such as for medical [13, 14] or military use cases [23]. Hence, the exploration of this unique collaboration could provide similar benefits in the high-stakes and complex work field of cybersecurity.

Due to its technical disposition, cybersecurity is a field where AI could be applied to a range of use cases; especially in the domain of detection and response, e.g., for continuous security monitoring [29, 49, 56, 68, 72, 101]. Other cybersecurity research is concerned with, e.g., the technical development of automated calculation of risk scores [76, 91], inferring the probability for a security incident [71], or dark web investigations for threat intelligence and text analysis [5, 28, 43].

However, not all cybersecurity tasks and decisions, such as assessments or stakeholder communication, can be easily automated. For example, tasks such as risk assessments require

USENIX Symposium on Usable Privacy and Security (SOUPS) 2024.
August 11–13, 2024, Philadelphia, PA, United States.

an understanding of the organisation's strategy and careful consideration of and communication with all stakeholders [44]. Hence, understanding the context, the experts' needs, and requirements for the intended collaboration is essential as it is influenced by the specific tasks, the required skills, and desired behaviours [79]. Furthermore, human-AI collaboration has often only been studied through the lens of novice users and non-specialised tasks; yet, experts have different requirements for working together with AI systems, as they, e.g., are more knowledgeable or self-confident in their domain of expertise and more averse to algorithmic advice [8, 12, 36, 70].

To provide a better understanding of the potential for expert-AI collaboration in cybersecurity, and the CSEs' requirements and perceptions, we conducted interviews with *N*=27 security professionals in Chief Information Security Officer (CISO) or related roles. As this research was exploratory with a focus on gaining an in-depth understanding, a qualitative interview approach using grounded theory was chosen.

The first objective of this research was to get a good understanding of security experts' tasks, and responsibilities, and to explore the potential CSEs see in collaborating with AI to complete their tasks, leading to the first research question (RQ).

RQ1: *What tasks and responsibilities do cybersecurity experts have, and which can be augmented by AI tools?*

We found that the experts' responsibilities are concerned with designing, implementing, and constantly reviewing and improving their organisations' cyber and information security strategies. The experts confirmed that, based on their assessments of the task nature, and both team members' capabilities, various tasks could be automated or augmented by AI.

After identifying potentially suitable tasks for collaboration, it was essential to understand the factors that are relevant for the successful collaboration of experts with AI in cybersecurity, e.g., the security experts' specific strengths, their requirements, and willingness related to the collaboration, and the capabilities they attribute to AI.

RQ2: *What is the cybersecurity experts' perspective on collaborating with AI tools to complete their tasks?*

The experts believed that AI can improve their workflows and relieve them of extensive and repetitive tasks, but also help with more discretionary tasks. Tasks such as monitoring, and analysing big amounts of data, or alerting the experts in unusual cases were identified as use cases where experts can benefit from the use of AI. On the contrary, especially stakeholder communication and the integration of the organisational context into the final decisions should remain in the human experts' domain.

Finally, user perceptions such as trust towards and concerns related to using AI tools have been shown to be relevant for the tool's acceptance and successful collaboration [18]. Of particular focus is how these might change across varying levels of autonomy and automation of AI tools concerning the level of transparency, autonomy, and adaptability of a system. AI tools can range from decision support tools, over collaborative scenarios where human approval is required, and situations where the human can only veto, to fully automated task execution by the AI.

RQ3: *What are the cybersecurity experts' perceptions on automation, autonomy, and trust in expert-AI collaboration?*

We found that deciding how autonomously an AI should be able to act was a difficult decision, which required experts to weigh factors such as their trust in AI tools and the suitability of the task for AI or the human expert, which influenced their assessment of deploying AI at different autonomy levels.

This research yielded valuable insights into the collaboration potential of AI and CSEs, which can guide the design of AI tools in cybersecurity. These tools can help fill the workforce gap by augmenting the existing CSEs' with AI, to free them up for other tasks that require their unique human capabilities and expertise. To that end, we provide two contributions:

First, within the high-stakes environment of cybersecurity, we explored for which type of tasks CSEs seek support from AI and outline how experts decided to share tasks between themselves and AI tools.

Second, we present a comprehensive autonomy decision framework that describes how the interplay of factors like the nature of the task, trust in AI, and a risk-benefit assessment impacts the decision to utilize AI on different autonomy levels. It provides a structured approach to determining the appropriate balance between human expertise and AI autonomy in cybersecurity.

These contributions help advance the understanding of expert-AI collaboration in cybersecurity and can guide the practical implementation of collaborative interfaces for CSEs and AI, fostering more effective and secure cyber defence strategies.

2 Related Work

In the following, we present related work concerned with AI in cybersecurity and expert-AI collaboration.

2.1 AI in Cybersecurity

Currently, the use of AI cybersecurity tools is relatively rare, with only one-third of organizations using or planning to use them [92]. Among these, 80% use AI to detect, 64% to predict, and 55% to reactively mitigate cyberattacks [92]. The primary applications for AI in cybersecurity are network security (75%), data security (71%), and endpoint security (68%) [15]. Diverse AI methods like machine learning (ML), or natural language processing (NLP) are promising for application in cybersecurity [77]. However, the use of AI also introduces novel risks, e.g., when ML models are tampered with during training through poisoning attacks, manipulating AI predictions [89]; or hallucinations, where LLMs produce syntactically correct answers that are nevertheless made up and false [6, 78].

Kaur et al. [46] identified literature concerned with the integration of AI into various use cases in cybersecurity and classified them according to the five NIST cybersecurity framework functions: identify, protect, detect, respond, and recover[1].

For the function of *identifying*, there are several studies concerned with supporting individuals and organisations to identify threats and risks, e.g., through automated calculation of risk scores [76, 91], or by inferring the probability for a security incident with AI [71]. The availability of related tools in practice is comparably mature: Tenable's Exposure AI [87], an attack surface management tool, or IBM's Guardium [39], which offers functionalities for risk assessment, vulnerability scanning and patch management. In the function of *protection*, AI can be used for threat simulation to identify and address gaps in software or misconfigurations in settings. When using AI in this function, e.g., for threat intelligence, it can efficiently combine data from multiple sources such as networks, users and IoT devices, for real-time monitoring and analysis [64]. Industry solutions supporting this function are IBM's Verify [40], offering AI functionalities for managing digital identities and access rights, and Zscalers Data Protection [102], providing data classification and visibility for the locations of sensitive data using AI. For the *detection* of anomalies, AI can increase event detection rates and detect unknown threats, as is demonstrated in AI tools for continuous security monitoring [29, 49, 56, 68, 72, 101]. Due to AI tools being able to monitor significantly more data, incidents can be reported more quickly and effectively [83]. Tools such as Microsoft's Security Copilot [62] support threat detection and prevention, and Tessian's Complete Cloud Security Email platform [88] uses AI to detect phishing and protect sensitive data on email. Once a cybersecurity incident is detected, AI tools can be used for the *response*, e.g., for automated isolation of affected entities [30, 74], or the automated remediation, such as preventing the spread of malware [38] or recommendations for countermeasures [67]. Industry solutions for the response function include Darktrace [22], which offers an automated response to threats such as ransomware, or Malwarebytes [55] and Kaspersky's Endpoint Security [45], offering AI-based malware identification and detection. The use of AI in the function of *recovery* is less mature, and the amount of research in this area is still relatively small compared to the other functions. Nevertheless, using AI for the aggregation of incidents [16], or the concluding analysis of vulnerabilities [60] are two examples of solutions in this function. Industry solutions for the recovery after a successful cyberattack are comparably scarce, however, Darktrace [22] offers an AI-based functionality that supports tasks in this domain, such as incident reporting.

Kaur et al. [46] emphasized the impact of human-AI collaboration in developing practical and usable AI for cybersecurity. Despite existing literature and AI tools from practice, the alignment with CSEs' needs, acceptance, or trust is rarely evaluated. Our research addresses this gap by examining experts' perspectives on AI adoption in cybersecurity, and identifying potential matches and mismatches with current tools and proposals.

2.2 Expert-AI Collaboration

Effective collaboration between humans and AI has been shown in research [25] and yet, the success of this collaboration depends on the understanding of delegation dynamics [8, 36, 70], is influenced by attitudes towards and knowledge of AI [70], and further factors such as overcoming cognitive biases [73] or algorithmic aversion which domain experts are especially prone to [12, 63, 100]. Supporting human experts with AI capabilities is based on the assumption that both actors can bring complementary skills to the collaboration, enhancing overall performance. While humans are suited for social tasks and unexpected situations, AI can perform repetitive and monotonous tasks quickly, accurately and reliably [35, 50]. In particular, AI can collect information logically and arithmetically and then process large amounts of data by weighting, prioritising, analysing and combining it [50]. In contrast, human actors can rely on their senses, emotional intelligence and social skills to build relationships and motivate employees [9, 23, 35]. Unlike AI, humans are able to use their intuition, creativity, and common sense in situations they were not trained for [35], enabling them to creatively develop solutions even in open and unfamiliar situations [48, 50, 94]. However, while existing research of human-AI collaboration yields insightful findings, they often stem from generalizable tasks performed by novices and might not hold in high-stakes environments or those that require expert knowledge as the domain of cybersecurity does [8, 36, 70]. Initial insights in the teaming of experts and AI show that especially in data-heavy fields, like medicine or military, where decisions are discretionary, rather than ruled by clear guidelines, expert-AI collaboration is beneficial to performance [2, 13, 24, 52]. While experts excel in unstructured environments, too many decision variables overwhelm human processing [3]. AI can reduce this complexity, enabling human experts to make informed decisions. Research on behaviours, skills, and abilities required for successful human-AI collaboration is fragmented and task-dependent [79], motivating our research to look specifically at the cybersecurity domain. As the context, the task, and the target group influence human-AI collaboration [79] we try to understand how CSEs and AI can collaborate across different

[1] https://www.nist.gov/cyberframework/getting-started/online-learning/five-functions

tasks. To do so, we introduce related frameworks describing relevant aspects of human-AI or related human-automation interaction in the following section.

2.3 Human-AI Collaboration Frameworks

Human-AI frameworks compare and describe interactions across various dimensions, such as trust [93], autonomy [37], or the influence of psychological factors [85]. For this purpose, Salikutluk et al. [75] evaluated the desired autonomy of a physical AI agent in a shared workspace, based on the situational context and the human agent's self-confidence, effects of task failure, understanding human capabilities (i.e., theory of mind), comparison of AI and human competence, and whether the human agent needs to adjust their actions. After refining the factors in a pre-study, they established that participants prefer an adapted autonomy level based on those factors, compared to fixed autonomy levels.

Similarly, Simmler and Frischknecht [80] derived a taxonomy describing human-AI collaboration that relies on two gradual parameters: *autonomy*, capturing the intractability of the system's actions, and *automation*, capturing the human level of control over the system's actions. The taxonomy defines levels of automation, reflecting the extent of human involvement in task execution and how autonomously an AI can act. At the lowest level, AI functions as a decision support tool, relieving experts of preparatory work for decision-making. Contrary, on the highest level five, there is no interaction between the human and the AI, with the AI being fully independent. Table 3 in Appendix A describes the five levels in more detail. The four key characteristics that define an autonomous system according to [80] include the system's transparency, determinism, adaptability, and openness (see Table 2 in Appendix A for a description). Overall, the taxonomy can be used to assess and describe the roles and responsibilities of each actor in human-AI collaboration, and therefore influenced the design of the AI autonomy section in our interviews.

The importance of trust for humans working with autonomous agents and the technology's acceptance has been shown in other research and motivated the trust section of our interviews [59, 81, 82, 97]. Research has shown that humans interact with machines differently than with other humans, nevertheless, there are similarities. For instance, humans apply human social rules and behaviour to machines [33, 34, 66], and trustor- and context-related factors have been considered equivalent regardless if the trustee is human or a machine [34]. In the organisational context, trust in technology describes the users' belief in the system's ability to perform as expected and the users' willingness to depend on the technology and make themselves vulnerable in uncertain and risky situations [31, 57, 65]. Theoretical descriptions of the human-technology trust relationship assume trust evolves over time; and, begins with an initial trust level even prior to the first interaction [65, 86]. This is then either confirmed or refuted upon the initial interaction with the technology [65, 86]. Succeeding the initial interaction, the users' trust was found to be positively related to the system's usability, and during the early periods also to the system's reliability [65]. When users do not understand the AI's prediction, and it conflicts with the implications of their own mental model, it can lead to uncertainties regarding decision-making [47]. One approach to support trust between humans and AI is the facilitation of transparency [32, 69, 97]. The provision of the AI's reasoning has shown to have a positive effect on cognition-based trust, indicating that this transparency can support bridging the discrepancies between the human mental model and the AI model and fostering the collaborative performance [97].

Therefore, through our interviews, we explored how experts would build trust in AI systems, and which factors could strengthen and weaken their trust.

Overall, the interplay of experts and AI needs to be better understood to enhance their collaboration, the expert's ability to evaluate the AI and its predictions, and the communication between the AI and the expert. Our study contributes to this understanding by exploring the CSEs' perspective on AI adoption in cybersecurity through in-depth interviews with CISOs and related job roles that additionally make use of existing frameworks such as the autonomy-automation taxonomy [80] to put the insights into a meaningful context.

3 Method

The following section describes the interview procedure, the data analysis procedure following a grounded theory approach, the sample and ethical considerations.

3.1 Interview Procedure

Through an online survey sent prior to the scheduled interview, the experts were informed about the data collection, processing, and storage and asked to give their consent to the explained procedures. To accompany the insights of the interviews, the survey contained a general attitude towards AI scale based on the *General Attitudes Towards Robots Scale (GAToRS)* [51], adapted to the AI-specific case and provided in Appendix C. Additionally, we asked experts to fill out the *Human-Computer Trust (HCT) scale* [54] in the pre-interview survey. The interviews were mainly conducted through Zoom or Microsoft Teams, with some in-person interviews. The interviews were audio-recorded after again obtaining verbal consent from the expert. The audio files were auto-transcribed with Trint [90], and then validated manually by two researchers. Data collection, transcription, and analysis occurred concurrently, in line with the grounded theory approach [21, 84].

The semi-structured interviews consisted of three sections. The full interview guide can be found in Appendix D.

1. The first section concerned understanding the CSEs' tasks, responsibilities, and the need for support.
2. The second section explored which type or level of expert-AI collaboration would be suitable for cybersecurity tasks, including a reflection on human and AI capabilities. This section was accompanied by an explanation of the automation levels [80] (detailed in Table 3). Experts went through their tasks and were asked to elaborate on the level of automation that was suitable for the integration of AI for each task. Then, CSEs were asked to sort the tasks into a matrix considering the feasibility and desirability of AI integration, similar to a How-Now-Wow matrix[2] as illustrated in Figure 2 in Appendix D.
3. The last part of the interview focused on understanding the CSEs' trust in and perceptions of AI tools, including their hopes and worries regarding expert-AI collaboration.

Interviews were internally piloted twice, to ensure the clarity of the questions, the feasibility of the methods and to approximate the interview duration. After piloting, the interview guideline was revised to reduce the number of questions, to avoid redundancy and, to change the order of questions in the section on AI and human capabilities. The final version can be found in Appendix D.

3.2 Data analysis

The survey data was analysed descriptively to contextualize the topic and obtain a comprehensive impression of CSEs' attitudes and trust towards AI.

To analyse the interview data, we followed a grounded theory approach [21, 84], which is suitable when little is known about the topic and allows synthesizing qualitative interview data and generating research assumptions and frameworks [19]. The interview guideline aimed to elicit diverse responses on the participants' perceptions initially, but provided guidance to experts, prompting them to evaluate AI in the context of specific cybersecurity tasks. Therefore, responses focused on similar tasks and allowed the emergence of consistent patterns.

We used the central element of grounded theory, ongoing memoing, in the transcription and during all coding phases, to capture impressions and ideas. Memoing describes the process of recording thoughts, analytical insights, decisions, and ideas in relation to the research process [21]. During the coding phase, we added the technique of diagramming [21], i.e., creating visual representations of interrelations between codes, to support the development of the categories and their relationships and interactions. The coding process was structured as follows: the initial open coding phase aimed at initial codebook development, where the first five interviews were coded with a line-by-line approach. Once an understanding of underlying themes in the data was developed, line-by-line codes were transformed into incidents.

The intermediate coding process focused on axial coding. Strauss and Corbin defined axial coding as *"a set of procedures whereby data are put back together in new ways after open coding by making connections between [and within] categories"* [21] (p.96). We captured the relationships of the arising themes and their contexts by diagramming and generating situational maps [20]. During axial coding, two researchers went through the interviews and developed a situational map depicting the codes until no new codes or relationships could be added.

In the final coding phase, the codebook derived from the situational maps was transferred to the coding software MAXQDA (v24.1.0) [95]. The interviews were then coded topic by topic, and interview sections discussing the same topics were compared between participants to enrich themes and provide different variations of one topic and respective codes. The full codebook can be found as an online appendix on https://doi.org/10.3929/ethz-b-000674517, and additional details are given in Appendix B.

3.3 Sample & Recruiting

The CSE sample was mainly recruited through purposive sampling to ensure that the participants matched the target group in terms of roles and level of expertise. Experts required a minimum of 2 years of industry experience in cybersecurity roles. As peer identification is another way to determine expertise, we used additional snowball sampling to help us recruit further relevant candidates for our interviews through interviewees and recruited one out of all 27 experts through this method. Experts were approached through social media platforms, specific expert forums, mailing lists, and peers. A total of 27 security experts from 23 different organisations were interviewed between November 2023 and January 2024, and their demographics are summarized in Table 1. The experts on average had M=15.37 years (SD=8.08) of experience in the field of cybersecurity. Throughout the rest of this paper, experts will be referred to as *ME*, when they are in a managerial role; *OE*, when they are in an operational role; and *CE*, when they are in a consultancy role. The individual years of experience for the experts can be found in Table 6 and the sample scores of the AI-adapted GAToR scale [51] are shown in Table 5 in Appendix C. Overall, on a 7-point Likert scale experts tended to agree with having personal experience with AI (M=5.50, SD=1.10), and being familiar with AI (M=5.54, SD=0.65). Additional information on the experts' organisations can be found in Appendix E.

[2] https://gamestorming.com/how-now-wow-matrix/

Gender		**Role**	
Male	25	Chief Information SO[3]	16
Female	2	Information SO[3]	2
Age		Chief SO[3]	2
25-34	4	Head of Security	2
35-44	7	Junior Information SO[3]	1
45-55	13	Security Architect	1
55-64	3	Network Security	1
HCT		Security Consulting Engineer	1
M=3.71 (*SD*=1.24)		Manager	1

Table 1: Expert demographics, $n = 27$

3.4 Ethical Considerations

Our institution's ethics board approved the study design, following established ethical guidelines for psychological research involving humans [7]. By collecting age ranges instead of a concrete age, we minimized the potential for privacy invasion. An informed consent form explaining the purpose of the study, data collection, and processing was given to participants before the interview. Participants were free to refuse participation and request the deletion of their data at any time without negative consequences. The audio data was transcribed, anonymized, and then deleted. The participants were given equal compensation and had the option of taking or donating the money to a charity of their choice.

4 Results

This section first describes the experts' responsibilities and tasks before summarizing the main themes that emerged in the interviews regarding the use of AI in cybersecurity, preferences for expert-AI task division, and trust in AI.

4.1 Cybersecurity Expert Roles, Responsibilities and Tasks

The experts described that their main responsibilities lie in the tactical and technical management of an organisation's information security, including the strategic information security orientation and ensuring adherence to regulatory and compliance requirements. The organisation's specific cyber and information security strategy is set out in policies, guidelines, and frameworks that are regularly reviewed, improved and audited. Experts mentioned responsibilities spanning from raising security awareness in the organisation to risk management, where threats to the organisation's assets are identified, evaluated, assessed, and appropriately treated. To protect the organisation from cyberattacks, experts described being responsible for technically steering the incident and vulnerability management, which includes planning, designing, and implementing technical systems, such as security information and event management (SIEM). Specifically, experts described communicating cybersecurity issues to management and other employees in an appropriate and sensitive manner, briefing them in the event of a crisis, and being available to answer ad-hoc questions. To ensure information security, experts also guide, plan and support the implementation of the organisation's security infrastructure. Expert responsibilities depend on their roles, where operational positions verify incidents, notify responsible parties, and facilitate device and entity recovery, bridging technical and human aspects within organisations. Experts in in-house managerial roles are primarily employed to fulfil the tasks described above, however inevitably get involved into more operational tasks if necessary. Experts in consultancy positions advise and consult external clients, and are less involved in the implementation and focus on the strategic and governance aspects, but held expertise in similar domains without hands-on involvement. Overall, the experts described the integration of AI into cybersecurity tasks as desirable, however, the tasks' suitability for AI integration varies based on AI and human capabilities.

Experts should plan, strategize and grasp the context. Tasks that participants considered lying in their responsibility often included competencies related to **strategy, planning, and assessment**. Experts considered these tasks to be more appropriate for humans, as "*transferring that to your context and to your company and to the risk that you have in your company. That's something that [AI] simply can't do that well.*" (CE24). Additionally, if there were no formal criteria that could be embedded into an AI to determine appropriate decisions or actions for specific scenarios, experts believed that humans should be responsible. Since those discretionary decisions would be "*something that has to do with emotion. It has to do with experience, it has to do with [...] circumstances and environmental influences. That's something that humans naturally take into account that a machine can't do.*"(ME5). For difficult discretionary decisions, with incomplete information, experts mentioned the need to weigh up equally valid options and in many cases relied not only on objective factors but also on experience from unrelated areas and their gut feeling. Typically, taking responsibility "*whether the activity, which can then be very drastic, makes sense in this context, this should still be a human who ultimately bears the responsibility for what happens.*" (ME8). While the integration of AI into these tasks was rated as somewhat desirable, most experts argued that this was not feasible at this time.

AI, the expert's little helper. However, this did not imply that the use of AI for these responsibilities was always deemed out of place. Many experts recognized the advantages of using AI to **prepare** tasks, such as assessments, audits, or guiding the development of strategies. Experts felt it was

[3] SO = Security Officer

suitable for AI to **gather information** on specific topics, **summarise** existing documents, and generate preliminary reports to guide the experts' decision-making. They recognised that AI could make vast amounts of information and data accessible to them through summaries. Leveraging AI's ability to bring "*information together and combine it in such a way that it makes some sense*" (ME20) to support decision-making during assessments was deemed to add value; "*[a]lthough a verification is then necessary.*" (ME20). Cybersecurity tasks in this category were gathering information about vulnerabilities, and incidents, giving insights into the current legal landscape or preparing information for risk assessments.

The age of generative AI. Experts also found the integration of AI useful for **generative** tasks. Most experts felt that letting AI write at least drafts for policies, frameworks, and guidelines would greatly support their work, but should be closely monitored and verified. They believed that AI would quickly learn the structure of such documents and could easily and efficiently reproduce them given different parameters to fit the respective organisation's needs. Additionally, experts mentioned that this would "*help me to create a proper language*" (ME7) for such compliance documents. The recent popularity and success of LLMs made them optimistic that AI "*will probably deliver a great policy*" (ME2), but also raised concerns that "*[AI] doesn't know my company, so it doesn't know the needs, demands, and requirements of my company*" (ME2). The experts were obviously aware of AI's fallibility, especially in regard to hallucinations, and insisted that no AI-generated document should just blindly be used but always be examined for context, and validity.

Can AI do awareness measures? Experts could not agree on how feasible or desirable the use of AI would be to raise cybersecurity **awareness**. Some experts believed that interpersonal interactions play a major role in the area of awareness, which makes the integration of AI counterproductive. Other experts elaborated on how they could use AI for raising awareness, "*for example, [to] produce an awareness training program tailored exactly to our needs: This is how I store data, make a short video about it, there's a quiz to review what has been learned*" (ME25). Some experts further mentioned that they were already using AI for awareness measures.

In summary, the human experts should be the ones driving the process, planning an awareness strategy and evaluating the achievement of their goals. AI could be utilized *within* the awareness measures, e.g., for generating and sending phishing emails to train employees, or generating content, such as texts, graphics, and videos for training purposes.

To communicate through AI or not to communicate through AI. Another task group that experts were reluctant to delegate was **communication**. While "*[AI] can make a lot of tasks easier, such as answering my emails, which I no longer want to do myself, and I am convinced that I could easily hand over to AI*" (ME27), other kinds of communication required the human experts' unique qualities. Communication, especially "*when it comes to crisis communication and so on, I would also say that you also need a good deal of empathy and political skill*" (ME23). Understanding and addressing the other party's needs through communication relies heavily on interpersonal aspects and sensitivity, which are rarely explicitly visible or graspable. Therefore, tasks such as crisis management and communication with customers or legal representatives were still considered best-suited for the human expert. The integration of AI was found to be undesirable and even infeasible.

However, answering employees' security-related questions was very desirable and feasible to delegate to AI. They pictured a compliance expert chatbot which would have access to their organisation's policies and implementations and thus be able to answer ad-hoc questions compliantly and tailored to the enquirer's role and level of expertise.

Leveraging AI capabilities to protect cybersecurity. The use of AI for **protection and prevention** was considered to make the experts' work easier and provide valuable support. While tasks like "*pen-testing requires a certain level of human intelligence, which is why it is so expensive and takes longer*" (ME20) , they expressed "*[i]f it were feasible, of course, [...] then it would be nice if it were as fully automated as possible. And then at the end you have a report [...]*" (ME20). In addition, AI could automatically review policy implementations or measures and identify potential gaps - "*tackle configuration management, verify firewall rules, check technical configuration elements*" (ME3) and potentially prevent successful attacks by "*enforcing requirements, [and] checking them*" (ME13).

The data-intensive and repetitive territory of AI. Tasks in the domain of **detection and response** were often mentioned as desirable and feasible for AI delegation. In particular, to "*analyse large volumes of data in a very short time, while incorporating historical data*" (ME4). In many cases, experts were aware that AI, compared to humans, can analyse data more precisely and differentiate benign and malignant patterns. In addition, experts noted that AI is not affected by human disadvantages: AI has no biorhythm and can therefore perform "*24/7/365 and at an alarming speed*" (ME18). Its impartiality and absence of emotions ensure consistent results and objectivity. Moreover, many experts were aware that an AI can "*reduce the sheer volume of data*" (ME8) and constantly "*learns more and more over time*" (ME2). The use of AI for monitoring could therefore replace and facilitate the work of many human analysts.

Although the experts generally viewed the use of AI for technical monitoring positively, monitoring humans raised ethical and legal concerns, making it undesirable. Letting the "*AI go there fully automatically and address or correct any misbehavior*" (ME2) was dismissed by experts.

The automated "*analysis for anomalies, performing baselin[ing] and outlier detection*" (ME9) was also found to be one of the most desired tasks for the integration of AI. However, the automatic response to detected anomalies, e.g., responding to suspected attacks by isolating entities, or automatically fixing detected vulnerabilities, involved more complex judgements. While the desirability of automatic responses was high, many experts were unsure about how much autonomy AI should have in these cases.

4.2 AI Autonomy in Cybersecurity

Experts were asked to elaborate on task division between them and AI, as well as the autonomy of AI in relation to the experts' previously described tasks. It is important to note that the concepts of AI automation and AI acting autonomously were used interchangeably during the experts' descriptions. As autonomy appeared to encompass aspects of automation, we will refer to the concept as "AI autonomy" from here on and detail the factors that experts deemed relevant for deciding on AI autonomy. The definition of autonomy as the "*quality or state of being self-governing*" [58] closely maps the experts' descriptions. Several experts were able to formally describe their internal thought process when evaluating AI autonomy levels for different tasks, as visualized in Figure 1. Deciding how autonomously AI should be able to act was a complex consideration for the experts. Factors such as the characteristics of the task, trust in AI, and risks and benefits played a role in this process. In the following sections, we first describe what factors experts evaluated regarding the task and their trust in AI. As shown in Figure 1, these two factors directly influence the experts' perceptions of the risk and benefit of different levels of AI autonomy. We will then describe the experts' assessments for the risks and benefits that influence what level of AI autonomy is deemed appropriate by CSEs (see Figure 1). This section concludes with the considerations for expert-AI task division at different levels of automation.

Considering the capability-task fit and the urgency. The nature of the task, and in particular how suitable experts **assessed** their **human capabilities** or **AI capabilities**, played into the decision for the AI autonomy level. For example, analytical tasks that require a lot of data processing were more likely to be assigned to AI with a greater level of autonomy, due to its efficiency and accuracy. In addition, the **urgency** of a task was also taken into account. Using AI at a high level of autonomy means it is able to carry out the work around the clock, which is important as "*the decision-making time and decision-making paths are becoming ever shorter. Time is becoming a massive success factor [...]*" (ME9).

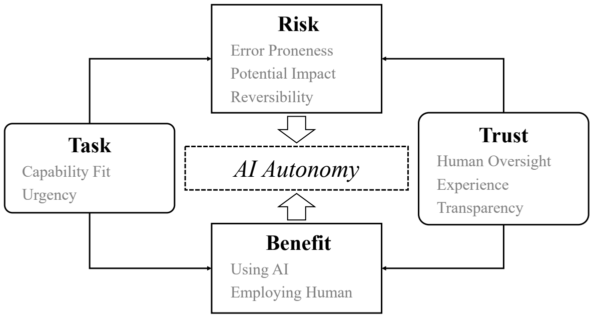

Figure 1: AI Autonomy Decision Framework

Trust: Experience and the importance of transparency. The experts' trust in AI played a major role in how autonomous they believed AI should be. We were able to identify three important aspects for trust in AI: human oversight, the experts' experience with AI, and AI transparency. A frequently articulated idea around **human oversight** was that "*in principle, I am fundamentally of the opinion that [AI] is a great support, but not with blind trust, never trust, and verify*" (ME8). This concept envisions an evolving process where humans closely monitor AI. However, as articulated by one expert, when "*you realize how good the output of the AI actually is*" (ME7), the necessity for human oversight diminishes. This shift is attributed to experts accumulating sufficient **experience** with AI, fostering a sense of trust. Additionally, comprehensive and sound regulation or trustworthy providers of AI models were mentioned as strengthening the trust in AI. Conversely, observing wrong AI predictions and experiencing misuse of AI for misinformation, propaganda, or social scoring, were detrimental to the expert's trust.

To better judge and incorporate AI results into their decisions, experts stressed the need for **transparent AI results** to understand the parameters leading to the AI's results, and to assess "*whether what the engine gives me is actually correct and has not been interpreted by the engine*" (ME12). Additionally, experts emphasized the importance of the **AI model's transparency** to understand how a model works, what data was used to train it, and what technical infrastructure it relies on. Uncertainties relating to the AI model's data processing and storage, or the potential accessibility of the data by malicious actors, were described as hindering the adoption by multiple experts.

Experts also stated a need for a deterministic tool that "*should always make the same decisions under the given parameters.*" (ME19), allowing them to rely on AI tools more. Overall, most experts expressed trusting AI, however many had conditions or requirements bound to their trust. Seven experts self-reported not to trust AI, as one put it "*I have relatively little trust in systems that act autonomously,[...] I really lack* *the basic trust that no unnecessary damage will be done.*" (ME9).

Risks: Error-proneness, impact and reversibility. AI's **proneness to error**, but also that of humans, played a central role in the experts' assessment of risks related to a certain level of autonomy. If AI were to be perfectly accurate in its calculations, predictions and detection rates, then experts felt it should be used for most cybersecurity tasks. Humans were described to be fallible and generally not suitable for many tasks in cybersecurity. Hence, AI without human influence could strengthen cybersecurity, as AI was portrayed as a reliable tool that is not influenced by emotions nor biased like humans might be.

Experts further determined the risk for a specific level of AI autonomy also by the respective **impact**, i.e., the extent and severity of the consequences, that an autonomous AI action could cause. When assessing consequences, experts considered how far-reaching they were, such as whether they were isolated to single workstations or could lead to company-wide shutdowns. The more severe and serious these consequences were perceived, the more conservatively experts were in granting higher levels of AI autonomy. Likewise, experts differed notably in their judgement based on task domain. For instance, for medical applications or critical infrastructure, experts were reluctant to consider allowing high AI autonomy.

In addition, the **reversibility** of a particular actions played a role. If the costs of a successful attack were high and the AI's action reversible, experts described that they would rather let AI react autonomous and too quickly and undo the action later than hesitate too long. "*Whether the impact of the action can be reversed and is fundamentally justifiable for the company*" (ME8) generally plays an important role in the allocation of autonomy. However, if the result is permanent, experts were more reluctant to grant AI autonomy.

Benefits: What can be achieved with AI, that can not be done without. When evaluating the benefits, experts compared **the advantages** that are **gained when AI is used** to advantages **gained when a human performs** a task. This assessment is predominantly determined by the nature of the task and the capability-task fit, as illustrated in Figure 1.

In the case that a task or an incident is time-sensitive, experts tended to lean towards giving AI more autonomy since "*[t]ime is becoming a massive success factor in [...], in a ransomware attack, the time between initial infection and outbreak is sometimes less than an hour. And people have to make sure they keep up.*" (ME9). They further believed "*that the reliability of an automated solution is generally higher.*" (ME5). In other words, this human approval can also be detrimental to the purpose.

Experts understood that "*if AI is perfectly trained to make a quick decision in a critical case, it may be better than a human being*" (CE24). Yet, experts also understood that the outcome and consequences of AI acting autonomously needs to be considered when deciding how much autonomy should be given to AI.

Level of autonomy: in between risks and benefits. Based on this framework of considerations (see Figure 1) some tasks were frequently mentioned for specific levels of autonomy and expert-AI task division. An additional tabular summary can be found in Appendix F.

Level 1 - Decision support: Particularly in areas with a high level of personal responsibility, or areas that require contextual knowledge or creativity, the experts stated that they would prefer to use AI only as a support for decision-making. This enables them to offload routine tasks partially, while potentially increasing their accuracy. In these cases, AI could summarise information to support risk management, or generate preliminary drafts for policies to support the experts in their tasks and processes.

Level 2 - Human Approval: When tasks fit the AI's capabilities, but are not time-critical but potentially have far-reaching consequences, experts increasingly expressed the wish to have AI propose a decision but still be able to review and ultimately accept or reject the proposed decision. Exemplar tasks mentioned by the experts were the response actions to major security incidents, including patching and the isolation of entities.

Level 3 - Human Veto: Regarding most tasks with far-reaching consequences or consequences that could not be easily reversed, the experts expressed a desire to at least be able to veto the actions of AI. While experts on the one hand wanted the option to intervene, AI could still automatically perform tasks without human interaction. However, experts seemed to have difficulty understanding, or applying this level to tasks that could be performed by AI.

Level 4 - Execute and Inform: In time-critical situations as well as in situations in which the consequences of the AI's actions could be mitigated or reversed, experts were willing to let AI autonomously execute tasks but still expressed the wish to be informed to maintain situational awareness. Exemplar tasks were automated penetration testing or the automated configuration of firewalls.

Level 5 - Fully automated: The experts mostly stated that they were willing to delegate routine tasks or tasks with minimal impact to AI for fully autonomous processing. In such cases, they would not want to be informed about each of the AI's actions, as this could quickly lead to information overload, especially in the context of daily routine tasks. Exemplar tasks were the distribution and verification of user privileges, and continuous security monitoring.

4.3 Designing Security Expert-AI Interaction

This section outlines the experts' requirements and preferences for designing expert-AI interfaces. One way in which experts described AI was a tool they could delegate time-consuming tasks to. A different, somewhat anthropomorphized view that the CSEs described was an AI assistant, where the AI's work is primarily supportive and at the request of the expert. Finally, several experts referred to AI as a co-pilot that actively thinks along and contributes to their shared work, as opposed to simply waiting to receive human input in order to perform tasks. Experts often considered themselves to be in a supervision role that evaluates, reviews and makes decisions with the support of AI. It was considered important for the AI tool to be available to the experts at all times and to be meaningfully integrated into their existing workflows. Experts emphasized the importance of seamlessly integrating AI into their existing processes, workflows, and routines so that day-to-day work is not complicated through additional steps. The communication between experts and AI needs to be appropriate to the situation. Many experts expressed the desire for communication through natural language. However, experts disagreed whether this should be through spoken word or via text. For important purposes, such as alerting, AI should address the expert proactively, whereas in other cases, experts wanted to be the ones initiating communication. The experts also urged that they must be able to integrate the context and strategy into the joint work when working on difficult cases. The experts therefore saw themselves in the role of the final decision-maker.

5 Discussion

In the following, we reflect on the limited deployment of AI tools in cybersecurity and factors that can enable successful integration. We discuss trust and its effects on experts' willingness to collaborate with AI and grant it specific levels of autonomy, the autonomy characteristics enabling expert-AI collaboration in cybersecurity. We conclude by highlighting the contributions of our AI autonomy framework for cybersecurity and provide design recommendations for expert-AI collaboration in cybersecurity.

Why is there little AI adoption? CSEs expressed wanting and needing support in various areas of their jobs, including generating policies and awareness materials, facilitating low-stakes communication, supporting compliance with regulations, as well as monitoring, pattern detection, and even automated threat responses. Our results show that experts would appreciate support from AI and are willing to widely employ it in cybersecurity, but that the use of AI for a majority of daily tasks is still perceived as hypothetical. Even though AI tools already exist for most cybersecurity use cases, as elaborated in subsection 2.1, only one-third of organisations reported using, or planning the use of AI tools for protecting their digital assets [46, 92].

One reason for this misalignment could be a mismatch between the characteristics of the currently available solutions and the experts' considerations for deploying AI with a certain level of autonomy. For example, our findings indicated that aside from the type of task, trust in AI played an important role in deciding on AI autonomy and adoption. As usage experience was mentioned as a relevant factor for trust in AI tools, another reason could be a lack of experience with existing AI tools and a potential hesitance to be among the first to adopt these tools in a high-stakes environment. The following sections discuss potential reasons along with implications related to building trust in AI tools that may ultimately enhance adoption rates of existing solutions or inform the development of matching AI tools.

The importance of good experiences and usability. The limited adoption of AI might also be related to negative experiences. Since experts stated to build trust primarily through experiences, encountering issues with AI systems can have a strong dissuading effect. This notion was reinforced by one expert, who reported a test with an AI tool in their organisation that turned out negatively, discouraging the expert from using AI in support of cybersecurity tasks. Other scientific studies confirm that if people can observe AI mistakes or malfunctions, the future outcomes of AI are viewed with more distrust [8], and its adoption becomes unlikely. Usability and reliability, at least in the early stages, impact the trust development with the introduced system [65], emphasizing the importance of these factors during system design. Also, in our interviews, experts placed a lot of importance on an initial familiarisation period. Conflicts during this initial exploratory phase could lead to an aversion to further, more sophisticated integration. AI tools, especially when used by domain experts, such as cybersecurity experts should be designed to be usable to that user group, and also prove to be reliable. Additionally, the quality of the AI tool should be carefully evaluated before deploying it, contributing to the secure integration of AI tools into an organisation, and at the same time also the trust building processes between experts and AI.

Building trust with the AI "employee". While experts almost unanimously described AI as a tool for their work, their descriptions of trust-building were closely related and sometimes even narrated using the example of a subordinate employee. In line with that, previous research has hypothesized that building trust between humans and AI seems to mirror processes of human-human relationships [34, 53]. This unintentional anthropomorphisation of the AI tools could indicate the experts' desire to relate with the AI tool and impact the experts' intention to use AI agents, but not directly induce trust [17, 98]. Experts repeatedly described that they would build trust in an AI tool by first giving it less important tasks, at a low level of autonomy, and closely monitoring its performance. Consistent with previous findings [1], experts appeared willing to give the system more freedom once the AI tool has proven to produce reliable results. Then, they might deploy a higher level of autonomy and use the AI for more significant tasks. Observing the trustee's performance, and their reliability were found to be correlated with the trustworthiness of a subordinate, and prior research already suggested that this might also apply to human-machine relations [34].

I should "never trust, and verify", or should I? Regardless of trust-building processes, for many experts trust in AI was still based on the possibility to oversee its work (see Figure 1). As experts have a high level of self-confidence and expertise in the domain where AI would be deployed, they could rely on their own capabilities to verify and correct AI results. This "never trust, and verify" principle, as one expert called it, strongly relies on the human as a final decision power to validate, verify and modify the AI outcome and to prevent damage or negative consequences. Especially, with the high availability of LLMs, the potential of AI to err has become even more visible to the user through AI hallucinations [6, 78], further increasing their desire to oversee the outcomes of AI models and validate their correctness. While manual verification is a good approach to mitigate negative consequences originating from the use of AI, it can also hinder the advantages gained through the collaboration of experts and AI. If experts always had the final say, AI systems would not only be inhibited in relieving experts of repetitive tasks due to requiring manual approval for actions but could even end up unable to react in real-time in critical situations. Lee and See [53] already pointed out that people with high self-confidence and low trust in automation tend to fall back to manual control more quickly, diminishing the benefits that AI could provide. To be able to leverage the capabilities of automation and intelligent systems in cybersecurity, it is therefore important that experts trust AI enough to the point where they do not feel a need to manually validate and verify all AI actions. Thus, the interaction of autonomy levels and the factors influencing trust need to be researched further to be able to design trustworthy and therefore effective AI systems for cybersecurity.

Influence of AI autonomy: considerations on AI adoption. At this point, we reflect on the experts' requirements for cybersecurity and how these relate to the four autonomy characteristics - adaptability, transparency, determinism, and openness - introduced by Simmler and Frischknecht [80] (see Table 2). First, the **adaptability** of the system was not so much a requirement for AI in cybersecurity expressed by the experts, but more so assumed by them to be one of the unique capabilities of AI as a technology. **Transparency** proved to be a relevant aspect influencing trust in AI, as shown in the AI autonomy decision framework in Figure 1. With the term transparency, experts mostly described the need for transparency and understandability of AI outputs, which is crucial for making quick yet well-founded decisions. The need for AI tools to be transparent arises from the complex and high-stakes nature of the cybersecurity environment. If discrepancies in judgment between the experts and AI arise, experts must be able to understand the factors that led to this discrepancy, allowing them to assess and correct the collaborative output based on their expertise. Similarly, Vössing et al. [97] argued that the ability to correct the output of AI, and exercise control can build trust. They found that providing explanations on the AI models' reasoning for collaborative task solving strengthened the cognition-based trust and reduced the discrepancies between the human mental model and the AI's embedded decision model, contributing to a successful collaboration [97]. Especially in the cybersecurity context, experts need to additionally understand the AI's reasoning to assess AI tools for potential tampering by malicious actors. Leveraging methods of explainable AI to provide transparency and information on the AI's reasoning therefore is a promising approach to improve collaboration of experts and AI in cybersecurity.

The need for transparency and sound reasoning in high-stakes decisions is not only preferred by experts but also a requirement by cybersecurity regulations [96], and an important factor in the EU AI Act for AI tools that humans interact with [27]. While the primary purpose of AI transparency was to gain trust, experts also described their need for results they can justify and comprehend in their responsibility towards management and legal authorities.

Further, experts need to be able to rely on AI tools that always react the same way given the same input for some critical cases, requiring a **deterministic** system. Such determinism can additionally increase the experts' perception of the AI's reliability, as it ensures predictability. Reliability additionally is an important aspect of trust [34], indicating that a semi-deterministic system could foster trust between experts and AI.

Many experts expressed worries about the potential to extract sensitive data if the AI tool was connected to additional sources or the internet, leading to the desire for transparency of AI models where they could at least partially quantify this risk and observe how the model handles, stores and distributes the data that users put in. An AI tool that could be deployed in cybersecurity needs to be **closed** [80] as the experts did not see the possibility of mitigating such risks otherwise. Crucially, this is important for even minor tasks like communication or summarising documents, as regulations and policies prohibit information from being shared. Therefore, any open system is not a viable option to enhance an organisation's cybersecurity, as the possibilities to tamper with AI models are still not yet fully understood nor mitigable.

In addition to the 'closed-loop' factor, it is also important to consider that feasibility also plays a role in the practical application of AI. Despite the digital nature of cybersecurity, most data still needs to be processed or even digitised in order to be useful for AI, which can further limit the practical feasibility for individual use cases. To wrap up our discussion and as a final conclusion, we compare our autonomy framework with existing frameworks in the literature. This should highlight differences, but also similarities, to show consistency with other literature on the one hand, but also uniqueness for the collaboration of cybersecurity experts with AI.

Comparison of autonomy-focused human-AI collaboration frameworks. While Salikutluk et al.'s [75] framework includes self-confidence as a fundamental factor, this factor is less relevant to our target population of CSEs, as it will always be high, and is therefore not represented in our framework. The effects of task failure and competence comparison are similar to our framework's risk and capability fit, respectively. As Salikutluk et al.'s framework draws from psychology, it includes theory of mind as an important factor that describes the understanding and awareness of the capabilities of different actors. This factor is included in our model through the notion of transparency, the CSEs' desire for transparent AI outcomes and models aligns with the need to understand and be aware of the other actor's capabilities [75]. While their model follows a flat structure, where all factors directly influence autonomy, our decision framework has two tiers, where the underlying task and trust factors then influence a risk and benefit trade-off. This difference might stem from CSEs having a more transactional view guided by higher-level cost versus reward considerations. CSEs are frequently exposed to risk-benefit analysis, and might naturally fall back to similar mechanisms for assessments of AI tools' autonomy. Our model is also not primarily developed under the assumption of a shared physical workspace setting; therefore, less emphasis is put on whether a change of human action is required. CSEs did emphasize the need for suitable integration of AI agents into their already existing workflows, making the change of human action undesirable. While both frameworks are established in different ways, ours through expert interviews imagining theoretical applications, and theirs with an emulated shared workspace with a physical AI agent, the similarities highlight the importance of some factors, shared between modalities. In particular, the capability fit for a task and the risk posed by failure seems to be perceived as important by humans when interacting with an AI system in either scenario. Future research should deepen the understanding of how the interaction medium and human expertise, and thereby self-confidence, affect the degree of autonomy afforded to AI systems.

5.1 Limitations & Future Work

Like all research, this study is subject to several limitations. *First,* our qualitative approach does not allow for quantification of the findings and can thus be viewed as an initial step towards informing future (quantitative) research on expert-AI collaboration in cybersecurity. *Second,* the sample was mostly male security professionals. This gender imbalance is not desired, but representative of the target group, as women are underrepresented in cybersecurity [42]. The sample included mainly experts with strategic and managing tasks. Future work could extend our high-level insights into cybersecurity professionals on operational levels. *Third,* we employed AI-adapted versions of the GAToRS [51] and the Human-Computer Trust scale [54] to better understand the participants' attitudes towards AI. However, it was challenging to identify meaningful quantitative patterns related to the multi-faceted qualitative data. Future work could evaluate these scales for AI-related research in quantitative settings with larger samples.

5.2 Summary: Recommendations for Expert-AI Collaboration in Cybersecurity

In sum, AI tools designed for effective collaboration with CSEs must address the demands of complexity, uncertainty, and high stakes in the field of cybersecurity. AI tools need to fulfil the high requirements related to data security and transparency and enable CSEs to make meaningful disclosures to management and legal authorities. At the same time, to leverage the complementary capabilities of experts and AI in cybersecurity, further understanding the effects of varying degrees of AI autonomy on expert trust is crucial for long-term adoption and successful collaboration. Therefore, the tool needs to have the following characteristics:

- the AI output needs to be transparent and understandable for CSEs, as well as the AI model and its respective infrastructure,
- the AI tool needs to be designed to accommodate the process of building trust allowing for low to high autonomy levels,
- the AI model needs to be closed to protect the data it is processing, semi-deterministic to accommodate for known best practices, but also adaptable to new threats to accommodate for the quickly evolving threat landscape of cybersecurity.

Data Availability Statement

Due to the high sensitivity of interviews with regard to the potential identification of participants through AI tools, the interview data is not openly available. Detailed sample information, the interview guide, codebook, and exemplary quotes are provided in the article and Appendix. For further information or access to the original interview transcripts, please contact the authors.

References

[1] Barbara D Adams, Lora E Bruyn, Sébastien Houde, Paul Angelopoulos, Kim Iwasa-Madge, and Carol Mc-

Cann. Trust in automated systems. Technical report, Ministry of National Defence, 2003.

[2] M.E. Ahsen, M.U.S. Ayvaci, and R. Mookerjee. When Machines Will Take Over? Algorithms for Human-Machine Collaborative Decision Making in Healthcare. In *Proceedings of the 56th Hawaii International Conference on System Sciences*, volume 202, pages 5733–5740, 2023.

[3] U. Aickelin, M. Maadi, and H.A. Khorshidi. Expert-Machine Collaborative Decision Making: We Need Healthy Competition. *IEEE Intelligent Systems*, 37(5):28–31, 2022.

[4] Matt Aiello, Scott Thompson, Max Randria, Camilla Reventlow, Guy Shaul, and Adam Vaughan. 2022 Global Chief Information Security Officer (CISO) Survey - Insights - Heidrick & Struggles. Technical report, Heidrick & Struggles, 2022.

[5] Khalid Al-Rowaily, Muhammad Abulaish, Nur Al-Hasan Haldar, and Majed Al-Rubaian. BiSAL – A bilingual sentiment analysis lexicon to analyze Dark Web forums for cyber security. *Digital Investigation*, 14:53–62, September 2015.

[6] Athaluri Sai Anirudh, Manthena Sandeep Varma, Kesapragada V. S. R. Krishna Manoj, Yarlagadda Vineel, Dave Tirth, and Duddumpudi Rama Tulasi Siri. Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*, 15(4), 2023.

[7] American Psychological Association et al. Ethical principles of psychologists and code of conduct. Technical report, American Psychological Association, 2016. Retrieved 14th February 2024 from `https://www.apa.org/ethics/code`.

[8] Elizabeth Bondi, Raphael Koster, Hannah Sheahan, Martin Chadwick, Yoram Bachrach, Taylan Cemgil, Ulrich Paquet, and Krishnamurthy Dvijotham. Role of Human-AI Interaction in Selective Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(5):5286–5294, June 2022.

[9] Adriana Braga and Robert K. Logan. The Emperor of Strong AI Has No Clothes: Limits to Artificial Intelligence. *Information*, 8(4):156, December 2017.

[10] Chuck Brooks. Cybersecurity Trends & Statistics For 2023; What You Need To Know, May 2023. Forbes.

[11] Chuck Brooks and Frederic Lemieux. Three Key Artificial Intelligence Applications For Cybersecurity by Chuck Brooks and Dr. Frederic Lemieux, September 2021. Forbes.

[12] Jason W. Burton, Mari-Klara Stein, and Tina Blegind Jensen. A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2):220–239, 2020.

[13] Federico Cabitza, Andrea Campagner, Luca Ronzio, Matteo Cameli, Giulia Elena Mandoli, Maria Concetta Pastore, Luca Maria Sconfienza, Duarte Folgado, Marilia Barandas, and Hugo Gamboa. Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis. *Artificial Intelligence In Medicine*, 138, April 2023.

[14] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), November 2019.

[15] Capgemini Research Institute. AI in Cybersecurity. `https://www.capgemini.com/news/press-releases/ai-in-cybersecurity/`, July 2019.

[16] Miguel V. Carriegos, Ángel L. Muñoz Castañeda, M. T. Trobajo, and Diego Asterio De Zaballa. On Aggregation and Prediction of Cybersecurity Incident Reports. *IEEE Access*, 9:102636–102648, 2021.

[17] Qian Qian Chen and Hyun Jung Park. How anthropomorphism affects trust in intelligent personal assistants. *Industrial Management & Data Systems*, 121(12):2722–2737, January 2021.

[18] Hyesun Choung, Prabu David, and Arun Ross. Trust in AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human–Computer Interaction*, 39(9):1727–1739, May 2023. arXiv:2203.12687 [cs].

[19] Ylona Chun Tie, Melanie Birks, and Karen Francis. Grounded theory research: A design framework for novice researchers. *SAGE Open Medicine*, 7, January 2019.

[20] Adele E Clarke. Grounded theory: Critiques, debates, and situational analysis. *The SAGE Handbook of Social Science Methodology*, pages 423–442, 2007.

[21] Juliet M. Corbin and Anselm Strauss. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1):3–21, March 1990.

[22] Darktrace. AI Cyber Security Solutions. `https://www.darktrace.com`, 2024. Accessed: 2024-06-06.

[23] Jack H Davenport. Collaborative human-machine analysis to disambiguate entities in unstructured text and structured datasets. In *Next-Generation Analyst IV*, volume 9851, pages 16–22. SPIE, 2016.

[24] Thomas H Davenport and Rajeev Ronanki. Artificial intelligence for the real world: Don't start with moon shots. *Harvard business review*, 96(1):108–116, 2018.

[25] Dominik Dellermann, Philipp Ebel, Matthias Söllner, and Jan Marco Leimeister. Hybrid Intelligence. *Business & Information Systems Engineering*, 61(5):637–643, October 2019.

[26] Mica R Endsley. The application of human factors to the development of expert systems for advanced cockpits. In *Proceedings of the Human Factors Society Annual Meeting*, volume 31, pages 1388–1392. SAGE Publications, Los Angeles, CA, 1987.

[27] European Union. Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206, 2021. COM/2021/206 final.

[28] Zhen Fang, Xinyi Zhao, Qiang Wei, Guoqing Chen, Yong Zhang, Chunxiao Xing, Weifeng Li, and Hsinchun Chen. Exploring key hackers and cybersecurity threats in Chinese hacker communities. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pages 13–18, September 2016.

[29] Alessandro Fausto, Giovanni Battista Gaggero, Fabio Patrone, Paola Girdinio, and Mario Marchese. Toward the Integration of Cyber and Physical Security Monitoring Systems for Critical Infrastructures. *Sensors*, 21(21):6970, January 2021.

[30] Lorenzo Fernández Maimó, Alberto Huertas Celdrán, Ángel L. Perales Gómez, Félix J. García Clemente, James Weimer, and Insup Lee. Intelligent and Dynamic Ransomware Spread Detection and Mitigation in Integrated Clinical Environments. *Sensors*, 19(5):1114, January 2019.

[31] David Gefen, Izak Benbasat, and Paula Pavlou. A Research Agenda for Trust in Online Environments. *Journal of Management Information Systems*, 24(4):275–286, April 2008.

[32] Shirley Gregor and Izak Benbasat. Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *MIS Quarterly*, 23(4):497–530, 1999. Publisher: Management Information Systems Research Center, University of Minnesota.

[33] P. A. Hancock. Politechnology: Manners Maketh Machine. *Human-Computer Etiquette: Cultural Expectations and the Design Implications They Place on Computers and Technology*, January 2010.

[34] P. A. Hancock, Theresa T. Kessler, Alexandra D. Kaplan, Kimberly Stowers, J. Christopher Brill, Deborah R. Billings, Kristin E. Schaefer, and James L. Szalma. How and why humans trust: A meta-analysis and elaborated model. *Frontiers in Psychology*, 14:1081086, March 2023.

[35] Patrick Hemmer, Max Schemmer, Michael Vössing, and Niklas Kühl. Human-AI Complementarity in Hybrid Intelligence Systems: A Structured Literature Review. In *PACIS 2021 Proceedings*, 2021.

[36] Patrick Hemmer, Monika Westphal, Max Schemmer, Sebastian Vetter, Michael Vössing, and Gerhard Satzger. Human-AI Collaboration: The Effect of AI Delegation on Human Task Performance and Task Satisfaction. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, IUI '23, pages 453–463, New York, NY, USA, March 2023. Association for Computing Machinery.

[37] Silvana Hinsen, Peter Hofmann, Jan Jöhnk, and Nils Urbach. How can organizations design purposeful human-AI interactions: A practical perspective from existing use cases and interviews. In *Hawaii International Conference on System Sciences*, 2022.

[38] Martin Husák. Towards a Data-Driven Recommender System for Handling Ransomware and Similar Incidents. In *2021 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 1–6, November 2021.

[39] IBM. IBM Security Guardium. https://www.ibm.com/security/data-security/guardium, Dec 2018. Accessed: 2024-06-06.

[40] IBM. IBM Security Verify. https://www.ibm.com/products/verify-saas, 2024. Accessed: 2024-06-06.

[41] ISACA. State of cybersecurity 2023 report, 2023. Retrieved 9th June 2024 from: https://www.isaca.org/resources/reports/state-of-cybersecurity-2023.

[42] (ISC)2. Cybersecurity Workforce Study, 2023. Retrieved 14th February 2024 from: https://www.isc2.org/research.

[43] Shintaro Ishikawa, Seiichi Ozawa, and Tao Ban. Port-Piece Embedding for Darknet Traffic Features and Clustering of Scan Attacks. In Haiqin Yang, Kitsuchart Pasupa, Andrew Chi-Sing Leung, James T. Kwok, Jonathan H. Chan, and Irwin King, editors, *Neural Information Processing*, Lecture Notes in Computer Science, pages 593–603, Cham, Switzerland, 2020. Springer International Publishing.

[44] Mohammad Hossein Jarrahi. Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4):577–586, July 2018.

[45] Kaspersky. Kaspersky Endpoint Security for Business. https://www.kaspersky.com/enterprise-security/endpoint, 2024. Accessed: 2024-06-06.

[46] Ramanpreet Kaur, Dušan Gabrijelčič, and Tomaž Klobučar. Artificial intelligence for cybersecurity: Literature review and future research directions. *Information Fusion*, 97:101804, September 2023.

[47] Ujwal Kayande, Arnaud De Bruyn, Gary L. Lilien, Arvind Rangaswamy, and Gerrit H. van Bruggen. How Incorporating Feedback Mechanisms in a DSS Affects DSS Evaluations. *Information Systems Research*, 20(4):527–546, December 2009.

[48] Gary A. Klein. *Sources of Power, 20th Anniversary Edition: How People Make Decisions*. MIT Press, September 2017.

[49] Hansaka Angel Dias Edirisinghe Kodituwakku, Alex Keller, and Jens Gregor. InSight2: A Modular Visual Analysis Platform for Network Situational Awareness in Large-Scale Networks. *Electronics*, 9(10):1747, October 2020. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[50] J. E. Korteling, G. C. van de Boer-Visschedijk, R. A. M. Blankendaal, R. C. Boonekamp, and A. R. Eikelboom. Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4, 2021.

[51] Mika Koverola, Anton Kunnari, Jukka Sundvall, and Michael Laakasuo. General Attitudes Towards Robots Scale (GAToRS): A New Instrument for Social Surveys. *International Journal of Social Robotics*, 14(7):1559–1581, September 2022.

[52] Chang-Eun Lee, Jaeuk Baek, Jeany Son, and Young-Guk Ha. Deep AI military staff: cooperative battlefield situation awareness for commander's decision making. *Journal Of Supercomputing*, 79(6):6040–6069, April 2023.

[53] John D. Lee and Katrina A. See. Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1):50–80, 2004.

[54] Maria Madsen and Shirley Gregor. Measuring human-computer trust. In *11th australasian conference on information systems*, volume 53, pages 6–8. Citeseer, 2000.

[55] Malwarebytes. Malwarebytes. https://www.malwarebytes.com, 2024. Accessed: 2024-06-06.

[56] Daniel L. Marino, Chathurika S. Wickramasinghe, Billy Tsouvalas, Craig Rieger, and Milos Manic. Data-Driven Correlation of Cyber and Physical Anomalies for Holistic System Health Monitoring. *IEEE Access*, 9:163138–163150, 2021.

[57] D. Harrison Mcknight, Michelle Carter, Jason Bennett Thatcher, and Paul F. Clay. Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, 2(2):12:1–12:25, July 2011.

[58] Merriam-Webster. autonomy. Retrieved 16th February 2024 from https://www.merriam-webster.com/dictionary/autonomy.

[59] Christian Meske and Enrico Bunde. Transparency and Trust in Human-AI-Interaction: The Role of Model-Agnostic Explanations in Computer Vision-Based Decision Support. In Helmut Degen and Lauren Reinerman-Jones, editors, *Artificial Intelligence in HCI*, volume 12217, pages 54–69. Springer International Publishing, Cham, 2020.

[60] Benjamin S. Meyers and Andrew Meneely. An Automated Post-Mortem Analysis of Vulnerability Relationships using Natural Language Word Embeddings. *Procedia Computer Science*, 184:953–958, January 2021.

[61] Microsoft. Microsoft erklärt: Was ist künstliche Intelligenz? Definition & Funktionen von AI, March 2020.

[62] Microsoft. Microsoft Security Copilot. https://securitycopilot.microsoft.com/, 2024. Accessed: 2024-06-06.

[63] Ali R. Montazemi. The impact of experience on the design of user interface. *International Journal of Man-Machine Studies*, 34(5):731–749, May 1991.

[64] Sridhar Muppidi, Lisa Fisher, and Gerald Parham. AI and automation for cybersecurity. Benchmark, IBM Institute for Business Value, United States of America, June 2022.

[65] Lea S. Müller, Christoph Nohe, Sebastian Reiners, Jörg Becker, and Guido Hertel. Adopting information systems at work: a longitudinal examination of trust dynamics, antecedents, and outcomes. *Behaviour & Information Technology*, 43(6):1096–1128, April 2024.

[66] Scott Brave Nass, Cliff. Emotion In Human-Computer Interaction. In *The Human-Computer Interaction Handbook*. CRC Press, 2 edition, 2007.

[67] Pantaleone Nespoli, Félix Gómez Mármol, and Jorge Maestre Vidal. A Bio-Inspired Reaction Against Cyberattacks: AIS-Powered Optimal Countermeasures Selection. *IEEE Access*, 9:60971–60996, 2021.

[68] Yannis Nikoloudakis, Ioannis Kefaloukos, Stylianos Klados, Spyros Panagiotakis, Evangelos Pallis, Charalabos Skianis, and Evangelos K. Markakis. Towards a Machine Learning Based Situational Awareness Framework for Cybersecurity: An SDN Implementation. *Sensors*, 21(14):4939, January 2021.

[69] Ingrid Nunes and Dietmar Jannach. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3):393–444, December 2017.

[70] Marc Pinski, Martin Adam, and Alexander Benlian. AI Knowledge: Improving AI Delegation through Human Enablement. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, pages 1–17, New York, NY, USA, April 2023. Association for Computing Machinery.

[71] Yuanqing Qin, Yuan Peng, Kaixing Huang, Chunjie Zhou, and Yu-Chu Tian. Association Analysis-Based Cybersecurity Risk Assessment for Industrial Control Systems. *IEEE Systems Journal*, 15(1):1423–1432, March 2021.

[72] Panagiotis Radoglou-Grammatikis, Panagiotis Sarigiannidis, Eider Iturbe, Erkuden Rios, Saturnino Martinez, Antonios Sarigiannidis, Georgios Eftathopoulos, Yannis Spyridis, Achilleas Sesis, Nikolaos Vakakis, Dimitrios Tzovaras, Emmanouil Kafetzakis, Ioannis Giannoulakis, Michalis Tzifas, Alkiviadis Giannakoulias, Michail Angelopoulos, and Francisco Ramos. SPEAR SIEM: A Security Information and Event Management system for the Smart Grid. *Computer Networks*, 193:108008, July 2021.

[73] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R. Varshney, Amit Dhurandhar, and Richard Tomsett. Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):83:1–83:22, April 2022.

[74] Jacob Sakhnini, Hadis Karimipour, Ali Dehghantanha, and Reza M. Parizi. Physical layer attack identification and localization in cyber–physical grid: An ensemble deep learning based approach. *Physical Communication*, 47:101394, August 2021.

[75] Vildan Salikutluk, Janik Schöpper, Franziska Herbert, Katrin Scheuermann, Eric Frodl, Dirk Balfanz, Frank Jäkel, and Dorothea Koert. An evaluation of situational autonomy for human-AI collaboration in a shared workspace setting. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–17, New York, NY, USA, May 2024. Association for Computing Machinery.

[76] José Carlos Sancho, Andrés Caro, Mar Ávila, and Alberto Bravo. New approach for threat classification and security risk estimations based on security event management. *Future Generation Computer Systems*, 113:488–505, December 2020.

[77] Iqbal H. Sarker, A. S. M. Kayes, Shahriar Badsha, Hamed Alqahtani, Paul Watters, and Alex Ng. Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, 7(1):41, July 2020.

[78] Yiqiu Shen, Laura Heacock, Jonathan Elias, Keith Hentel, Beatriu Reig, George Shih, and Linda Moy. ChatGPT and Other Large Language Models Are Double-edged Swords. *Radiology*, 307(2):e230163, April 2023.

[79] Dominik Siemon. Elaborating team roles for artificial intelligence-based teammates in human-ai collaboration. *Group Decision and Negotiation*, 31(5):871–912, 2022.

[80] Monika Simmler and Ruth Frischknecht. A taxonomy of human–machine collaboration: capturing automation and technical autonomy. *AI & SOCIETY*, 36(1):239–250, March 2021.

[81] Matthias Söllner, Axel Hoffmann, Holger Hoffmann, Arno Wacker, and Jan Marco Leimeister. Understanding the Formation of Trust in IT Artifacts. In *Proceedings of the International Conference on Information Systems, ICIS 2012*, volume 11, Orlando, December 2012. Association for Information Systems.

[82] Matthias Söllner, Axel Hoffmann, and Jan Marco Leimeister. Why different trust relationships matter for information systems users. *European Journal of Information Systems*, 25(3):274–287, May 2016.

[83] Richard Starnes, Sumit Cherian, and Luis Delabarre. Reinventing cybersecurity with artificial intelligence: The new frontier in digital security, July 2019. Retrieved 16th February 2024 from `https://www.capgemini.com/insights/research-library/`.

[84] Anselm L. Strauss and Juliet M. Corbin. *Basics of qualitative research: techniques and procedures for developing grounded theory*. Sage Publications, Thousand Oaks, 2. edition, 2003.

[85] S Shyam Sundar. Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication*, 25(1):74–88, March 2020.

[86] Matthias Söllner and Paul Pavlou. A longitudinal perspective on trust in it artefacts. *Research Papers*, June 2016.

[87] Tenable, Inc. Tenable One Exposure Management Platform. https://www.tenable.com/products/tenable-one, October 2022. Accessed: 2024-06-06.

[88] Tessian. Complete Cloud Security Email Platform. https://www.tessian.com, 2024. Accessed: 2024-06-06.

[89] Zhiyi Tian, Lei Cui, Jie Liang, and Shui Yu. A Comprehensive Survey on Poisoning Attacks and Countermeasures in Machine Learning. *ACM Computing Surveys*, 55(8):166:1–166:35, December 2022.

[90] Trint Limited. Trint. https://trint.com/, 2022.

[91] Agnieszka A. Tubis, Sylwia Werbińska-Wojciechowska, Mateusz Góralczyk, Adam Wróblewski, and Bartłomiej Ziętek. Cyber-Attacks Risk Analysis Method for Different Levels of Automation of Mining Processes in Mines Based on Fuzzy Theory Use. *Sensors*, 20(24):7210, January 2020.

[92] TÜV-Verband. Einsatz künstlicher Intelligenz in der IT-Sicherheit in deutschen Unternehmen 2020. https://de.statista.com/statistik/daten/studie/1251115/umfrage/ki-in-der-it-sicherheit-in-unternehmen/, 2020.

[93] Takane Ueno, Yuto Sawa, Yeongdae Kim, Jacqueline Urakami, Hiroki Oura, and Katie Seaborn. Trust in human-AI interaction: Scoping out models, measures, and methods. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7. ACM, April 2022.

[94] Erik Veitch and Ole Andreas Alsos. A systematic review of human-AI interaction in autonomous ship systems. *Safety Science*, 152:105778, August 2022.

[95] VERBI Software. MAXQDA, 2024. https://www.maxqda.com/.

[96] Javier Verdugo and Moisés Rodríguez. Assessing data cybersecurity using ISO/IEC 25012. *Software Quality Journal*, 28(3):965–985, September 2020.

[97] Michael Vössing, Niklas Kühl, Matteo Lind, and Gerhard Satzger. Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*, 24(3):877–895, June 2022.

[98] Katja Wagner, Frederic Nimmermann, and Hanna Schramm-Klein. Is It Human? The Role of Anthropomorphism as a Driver for the Successful Acceptance of Digital Voice Assistants. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[99] Johannes Weyer. Die Kooperation menschlicher Akteure und nicht-menschlicher Agenten: Ansatzpunkte einer Soziologie hybrider Systeme. Technical report, Wirtschafts- und Sozialwissenschaftliche Fakultät Universität Dortmund, 2006. Working Paper.

[100] Stacey M Whitecotton. The effects of experience and confidence on decision aid reliance: A causal model. *Behavioral Research in Accounting*, 8:194–216, 1996.

[101] Fan Zhang, Hansaka Angel Dias Edirisinghe Kodituwakku, J. Wesley Hines, and Jamie Coble. Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data. *IEEE Transactions on Industrial Informatics*, 15(7):4362–4369, July 2019.

[102] Zscaler, Inc. Zscaler Data Protection. https://www.zscaler.com/solutions/security-transformation/data-protection, 2024. Accessed: 2024-06-06.

Appendix

Appendix A: Automation/Autonomy Taxonomy

The following tables 3 and 2 detail the levels of automation and the dimensions of autonomy as described by [80].

Dimension	Description
Transparency	Degree to which all execution steps between an input A and an output Be are specified and transparent
Determinism	Degree to which an A always equally leads to an output B
Adaptability	Degree to which a system can learn and adapt behaviour to changing environments
Openness	Degree to which the system can expand its original input for collaboration and interaction

Table 2: Dimensions of autonomy based on Simmler & Frischknecht [80].

Level	Description	Explanation
1	Offers Decision	System makes recommendations, operator selects and decides
2	Executes with human approval	System makes recommendations and selects "best" option, operator (dis-)approves
3	Executes if no human vetoes	System makes recommendation, selects "best" option and executes, operator can correct and veto
4	Executes and then informs	System makes recommendation, selects "best" option, executes, and informs operator (passive operator role)
5	Executes fully automated	System makes recommendation, selects "best" option and executes without informing (operator not part of process)

Table 3: Levels of automation as described by Simmler & Frischknecht [80] and based on Endsley [26] and Weyer [99].

Appendix B: Complete Codebook

The complete codebook, including descriptions and examples for each code, have been published on the ETH Research Collection and are accessible via the following DOI: https://doi.org/10.3929/ethz-b-000674517.

Appendix C: AI-adapted General Attitudes Towards Robots Scale (GAToRS) [51]

No.	AI-adapted Item description
Personal Level Positive Attitude (P+)	
RA1	I can trust persons and organizations related to development of AI
RA2	Persons and organizations related to development of AI will consider the needs, thoughts and feelings of their users
RA3	I can trust in AI
RA4	I would feel relaxed interacting with an AI
RA5	If AI had emotions, I would be able to befriend them
Personal Level Negative Attitude (P-)	
RA6	I would feel uneasy if I was given a job where I had to use AI
RA7	I fear that an AI would not understand my commands
RA8	AI scares me
RA9	I would feel very nervous just being around an AI
RA10	I don't want an AI to talk to me
Societal Level Positive Attitude (S+)	
RA11	AI is necessary because it can do jobs that are too hard or too dangerous for people
RA12	AIs can make life easier
RA13	Assigning routine tasks to AIs lets people do more meaningful tasks
RA14	Dangerous tasks should primarily be given to AI
RA15	AI is a good thing for society because it helps people
Societal Level Negative Attitude (S-)	
RA16	AI may make us even lazier
RA17	Widespread use of AI is going to take away jobs from people
RA18	I am afraid that AI will encourage less interaction between humans
RA19	AI is one of the areas of technology that needs to be closely monitored
RA20	Unregulated use of AI can lead to societal upheavals
Criterion Items	
C1	Generally speaking, I have a positive view of AI
C2	I have personal experience of using AI
C3	I am interested in scientific discoveries and technological developments
C4	AI is a familiar topic to me

Table 4: GAToRS by [51] and adapted to the AI use case.

Sub Scale	Min	Mean	Max	SD
Personal+	13.00	19.81	24.00	3.12
Personal-	7.00	12.19	19.00	3.01
Societal+	18.00	25.15	31.00	3.06
Societal-	11.00	23.19	28.00	4.11

Table 5: Sample scores of the AI-adapted GAToR Scale [51], $n = 27$

Appendix D: Interview guideline

The following section describes the interview guideline consisting of three focus areas: 1) Understanding the job, 2) Understanding the type or level of human-AI collaboration, and 3) Understanding the hopes, fears, and emotions.

Focus 1: Understanding the job.
RQ: What tasks do cybersecurity experts need to complete, and which could hypothetically be automated or complemented by AI?
Preface: Talk about role in general, not considering AI at this point.
Role description
What is your current role in your organization? What are your key responsibilities in your role? Ask if not already mentioned: Which tasks do you need to do in order to achieve your goal(s) and fulfil your key responsibilities?
Need for support
For which of the tasks are you lacking resources, e.g., time or skills? In case not answered: Which tasks would you like support from an automation or AI solution for?
Focus 2: Understanding the type or level of human-AI collaboration.
RQ: What is the expert's view on the feasibility of integrating AI into their workflow to collaborate?
Reflecting on Human Capabilities
What do you imagine AI is generally good at doing? Which of the tasks are well suited for AI in the domain of cybersecurity?
Reflecting on AI Capabilities
What do you imagine AI is generally good at doing? Which of the tasks are well suited for AI in the domain of cybersecurity?
Collaboration of Humans and AI
What would the interaction between you as an expert and the AI be?
Feasibility
We have talked about which tasks are well suited, but for which would you like to integrate AI and why? Where would you hesitate to use AI? What kind of support do you require? E.g., decision-aid (i.e., information gathering) or automated responses (i.e., AI can react to detected threats itself) etc. Why?
Feasibility *(continued)*
What limitations might arise from AI as a technology? In practice, what might be other factors limiting or impacting the feasibility of the human-AI collaboration?
Having discussed the tasks and aspects of their feasibility now, I would like to ask you to sort the tasks into this how-now-wow matrix (visualized in Figure 2). The tasks and collaboration ideas you think feasible and find somewhat desirable should go into now. The tasks and collaboration ideas you find desirable but are unsure of the feasibility should go into the how, and ideas that are highly desirable and at the same time highly feasible, please place into the wow matrix.
Focus 3: Understanding the hopes, fears, and emotions.
(Former) RQ: Which emotions, hopes, and fears do expert-AI collaboration trigger in experts, and for which reasons?
Trust
Could you, as one half of this collaboration, trust the AI? What do you need to establish and maintain trust in this technology and the collaboration? What could cause you to lose trust in AI?
Hopes
What are your hopes considering expert-AI collaboration?
Worries
What are your worries about the integration of AI into your workflow?

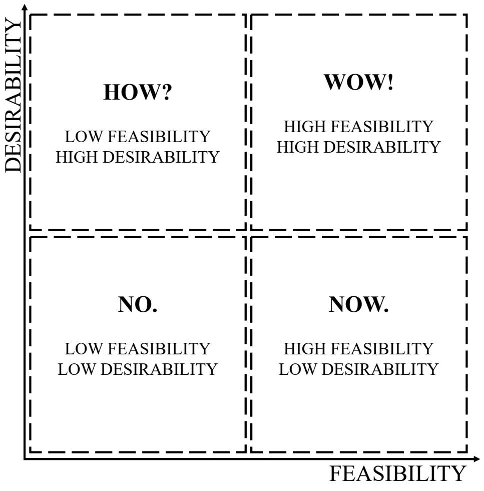

Figure 2: Feasibility-Desirability Matrix

Appendix E: Detailed Demographics

The following tables detail the expert demographics in terms of work experience, and the characteristics of the experts' organizations.

Expert	Experience (Years)	Role	Gender
ME1	2	Junior Information Security Officer	f
ME2	10	Information Security Officer	m
ME3	25	Chief Information Security Officer	m
ME4	20	Chief Information Security Officer	m
ME5	15	Chief Information Security Officer	m
ME6	35	Chief Information Security Officer	m
ME7	22	Former Chief Information Security Officer, now: Founder & Owner	m
ME8	23	Head of Security	m
ME9	15	Security Architect	m
ME10	5	Head of Security	m
OE11	18	Network Security	m
ME12	20	Chief Security Officer	m
ME13	10	Chief Security Officer	m
ME14	25	Chief Information Security Officer	m
ME15	25	Chief Information Security Officer	m
ME16	20	Information Security Officer	m
ME17	15	Chief Information Security Officer	m
ME18	14	Chief Information Security Officer	m
ME19	4	Chief Information Security Officer	m
ME20	13	Chief Information Security Officer	m
ME21	15	Chief Information Security Officer	m
ME22	12	Chief Information Security Officer and Data Protection Officer	m
ME23	20	Chief Information Security Officer	m
CE24	3	Security Consulting Engineer	f
ME25	15	Chief Information Security Officer	m
ME26	4	Chief Information Security Officer	m
ME27	10	Manager	m

Table 6: Expert description by years of experience and role

# employees	Count
<50	2
51-200	1
201-500	3
501-1000	3
1001-5000	3
5001-10000	6
>10000	6

Table 7: Experts organisations size, n=24

Industry	Count
Banking, Finance and Insurance	6
Consulting	1
Education	1
Health Services	1
IT Services	5
Manufacturing	1
Marketing	1
Media	1
Public Services	3
Telecommunication	1
Transport	2
Utilities	1
Country	**Count**
America	1
Germany	1
Switzerland	22

Table 8: Experts organisations industry and location, n=24

Appendix F: Autonomy Levels and Tasks

LVL	Requirements for Tasks	Task Examples
1	Personal responsibility, require contextual- or goal-understanding, or creativity	Risk management and assessment, policy development, solution architecture development
2	AI capabilities fit, non-time-critical, potentially far-reaching consequences	Patching vulnerabilities, isolating infected assets after incidents, generating content for trainings
3	Far-reaching and non-reversible consequences	Storage management
4	Time-critical and reversible, consequences could be mitigated, maintain situational awareness	Vulnerability detection, firewall configuration, monitoring network traffic, log-file analysis, detection of phishing emails
5	Routine and minimal consequences, avoid information overload	Distribute and verify user privileges, simulating phishing emails for training, malware detection

Table 9: Autonomy Levels with Tasks