Computers in Human Behavior 138 (2023) 107451

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

# Adapt and overcome: Perceptions of adaptive autonomous agents for human-AI teaming

Allyson I. Hauptman *, Beau G. Schelble, Nathan J. McNeese, Kapil Chalil Madathil

*Clemson University, P.O. Box 1212, Clemson, SC, 29632, USA*

ARTICLE INFO

*Keywords:*
Autonomous agents
Adaptive autonomy
Human-autonomy teaming
Cyber security
Computer security
Incident response

ABSTRACT

Rapid advances in AI technologies have caused teams to explore the use of AI agents as full, active members of the team. The complex environments that teams occupy require human team members to constantly adapt their behaviors, and thus the ability of AI teammates to similarly adapt to changing situations significantly enhances the team's chances to succeed. In order to design such agents, it is important that we understand not only how to identify the amount of autonomous control AI agents have over their decisions, but also how changes to this control cognitively affects the rest of the team. Professional organizations often break their work cycles into phases that set limits on the team members' actions, and we propose that a similar process could be used to define the autonomy levels of AI teammates. Cyber incident response is an ideal context for this proposal, as we were able to use incident response phases to explore how a team's work cycle could guide an AI agent's changing level of autonomy. Using a mixed methods approach, we recruited 103 participants to complete a factorial survey containing ten contextual vignettes focused on an AI teammate's level of autonomy in incident response contexts, and from these participants we conducted twenty-two follow-on qualitative interviews that further explored how the participants felt an AI agent's adaptive capabilities would affect team performance and cohesiveness. Our results showed that work cycles can be used to assign autonomy levels to adaptive AI agents based upon the degree of formal processes and predictability of the team's tasks during the cycle, and that dynamic, human-like adaptation methods are vital to effective human-AI teams. This research provides significant contributions to the HCI community by proposing design recommendations for the development of adaptive autonomous teammates that both enhance Human-AI teams' productivity and promote positive team dynamics.

## 1. Introduction

Exponential growth in the area of artificial intelligence (AI) has led to new frontiers of interaction between humans and AI agents (O'Neill, McNeese, Barron, & Schelble, 2020). One area in which these increasing levels of interaction can be both problematic and beneficial is human-AI teaming, where AI agents collaborate with humans as interdependent teammates to reach a common goal (McNeese, Demir, Cooke, & Myers, 2018). Teams operate in dynamic environments that force them to adapt their role over time and offload or take on additional tasks based on their cognitive load and environmental constraints (de Greef & Arciszewski, 2008). AI teammates pose a unique challenge to teams, because AI designers must determine the amount of autonomy these agents should possess. The amount of autonomy that an AI agent possesses can be described as its level of autonomy (LOA) (O'Neill et al., 2020), a scale derived from the levels of automation (Parasuraman, Sheridan, & Wickens, 2000). It has been proposed that AI agents could adapt their programmed level of autonomy, a concept referred to as adaptive autonomy (Suzanne Barber, Goel, & Martin, 2000); however, adaptive AI agents have not been studied within human-AI teaming contexts.

Adaptive AI agents present an additional challenge to human-AI teaming, as human teammates must repeatedly adjust their interactions and information sharing requirements with the agent as its level of autonomy changes (Tambe, Pynadath, Chauvat, Das, & Kaminka, 2000). This means that an AI agent that conducts multiple adjustments to its capabilities, in this case its level of autonomy, might cause multiple changes to the team's shared awareness in the pursuit of its common goals. This could be an issue, as teammates rely upon shared awareness in order to understand and predict each other's team roles and actions (Smith, 2019). Previous research has shown that

* Corresponding author.
*E-mail addresses:* ahauptm@g.clemson.edu (A.I. Hauptman), bschelb@clemson.edu (B.G. Schelble), mcneese@clemson.edu (N.J. McNeese), kmadith@clemson.edu (K.C. Madathil).

https://doi.org/10.1016/j.chb.2022.107451
Received 9 March 2022; Received in revised form 6 August 2022; Accepted 21 August 2022
Available online 26 August 2022

because humans and AI agents do not communicate in the same ways, this shared awareness is already difficult to maintain in human-AI teams (McNeese, Schelble, Canonico, & Demir, 2021). For this reason, in addition to examining professional perceptions of the right levels of autonomy for an adaptive AI teammate, we chose to also research how the team's perceptions of the AI teammate might change as it adapted. In particular, we wondered how the human teammate's perceptions of the AI agent would be affected by the ability to control the agent's autonomy level.

While there have been a number of recent research studies on human-AI teams in gaming and piloting contexts (Liu, Lai & Tan, 2021; Zhang, McNeese, Freeman, & Musick, 2021) little research has focused on the role of AI teammates for other professional organizations. Professional teams add an additional layer of complexity, as these teams continuously receive tasks and restraints from external sources, such as legal dictates and third party clients, that bear significant real-world consequences (Staves, Balderstone, Green, Gouglidis, & Hutchison, 2020). For instance, failing to adhere to a client's requirements may cost a team the job, and thus the compensation and employment that comes with it. In fact, many technical professions like incident response (Mitropoulos, Patsos, & Douligeris, 2006), software development (Jain & Suman, 2015), and cyber operations (Wen, Rao, & Yan, 2018), purposefully break their work cycles into distinct phases or tasks so that teams understand what actions are relevant and acceptable during that defined period (Staves et al., 2020). In this way, human professionals constantly adapt their level of autonomy to comply with the constraints of the situation over time. In other words, we learn through training and experience what the left and right limits are for working independently and when we need the permission and guidance of a higher authority. Quite a few professions, particularly those associated with technical domains, are seeking to increase the use of AI agents as a solution to current expertise and capability gaps (Nyre-Yu, Gutzwiller, & Caldwell, 2019). In particular, these organizations are considering how to utilize adaptive AI agents, which would be capable of altering their level of autonomy over the course of a defined task (de Greef & Arciszewski, 2008). Thus, a professional team that already has a defined methodology that dictates the autonomy of human teammates serves as an excellent context in which to study the implementation of adaptive AI teammates.

Cyber incident response teams are an ideal case study for this research. Incident response is a fast-paced, complex field in which a variety of experts must work together to prevent, identify, and correct security incidents (Nyre-Yu, 2019). As technology increases in use and sophistication, so do the attacks that target it, and these challenges require advanced prevention, detection and response techniques to counter (Donevski & Zia, 2018). AI agents have emerged as a solution to some of these problems, as they are capable of quickly analyzing large data sets far beyond human capability, which has made them attractive to a multitude of organizations seeking to automate network and computer security roles (Burke, 2020). These AI agents could undoubtedly increase the capabilities of incident response teams, but there is a distinct lack of research into the current and possible use of AI agents in computer security contexts (Jarrett & Choo, 0000), which makes the questions that this paper asks not only important HCI research, but to computer security research, as well.

The outlined research gaps are addressed by the two following research questions, each divided into two sub-questions. These research questions directly address the practical implementation of adaptive AI teammates according to a team's work cycle, and the effects of that implementation on the rest of the team's perceptions of their AI teammate.

**RQ1**: What are perceived optimal levels of autonomy for an adaptive teammate throughout the phases of a Human-AI team's work cycle?

*RQ 1.1*: Can team members agree what level of autonomy best meets the needs of each phase?

*RQ 1.2*: What are the characteristics of each phase that dictate the team's comfort with higher or lower levels of autonomy?

**RQ2**: Does having the ability to adjust an AI agent's level of autonomy alter its human teammates' perceptions of it as a legitimate teammate?

*RQ 2.1*: Does it create an improper power balance amongst teammates?

*RQ 2.2*: Does whether the agent adapts manually or dynamically change these perceptions?

This paper utilizes a mixed methods approach to investigate our questions on the optimal levels of autonomy and effects of adaptive autonomous agents on their teammates' perceptions. Using a series of computer security incident vignettes and the well-defined incident response cycle, it will show how a professional work cycle that teams use to dictate human actions can be used to define the appropriate autonomy levels and adaptation points of AI teammates. AI agents must not only adapt to their environment, but to the changing needs and comfort levels of their human teammates. This research is incredibly important for all organizations seeking to utilize adaptive autonomous teammates. As a diverse, complex team of technical, legal, and political professionals (Nyre-Yu et al., 2019) the processes and perspectives of an incident response team can be extrapolated to fit a number of professional human-AI teaming contexts. Indeed, the possibilities for AI on these teams go beyond even software-defined AI to include a wide array of embodied agents that can be used to work and communicate with humans (Rist et al., 2004), and thus render it even more applicable to the wider community. The answers to RQ1 and RQ2 also provide the community with important design implications for incorporating AI teammates into professional teams and the subsequent effects those adaptive AI agents have on their human teammates and the overall team dynamic.

## 2. Related work

### 2.1. Human-AI teaming

The way in which humans behave on teams fundamentally changes with the addition of AI teammates (Demir & Cooke, 2014). A recent review of the empirical human-AI teaming literature defined a human-AI team as consisting of at least one human and one artificial agent in which the teammates are interdependent, share a common goal, and possess a significant degree of independence (O'Neill et al., 2020). The idea that AI agents could interact with humans with such heightened intelligence is due to recent advances in Deep Learning technologies that have significantly broadened the horizons of AI agents (Oh et al., 2018). Previous research has shown that numerous team situations can substantially benefit from an AI teammate, such as data science (Wang et al., 2019) and computer security (Mahaini, Li, & Sağlam, 2019), however; team performance can differ based on whether the AI is viewed more as a tool or legitimate teammate (Zhang et al., 2021). There are a variety of factors that could influence how human teammates perceive AI teammates, including predictability, directability, and common ground (Klien, Woods, Bradshaw, Hoffman, & Feltovich, 2004). Recent research has also shown that the most important factor that influences a human teammate's perception of an AI agent is the AI's skill level; in particular, humans were far more receptive and positive towards working with an AI teammate if they felt it had high level, relevant skills (Zhang et al., 2021). Some of this could be due to previous experience with rudimentary or faulty AI, experiences which fuel skepticism of an AI teammate's skills. This is why the first question posed by this study focuses on the AI agent's ability to act autonomously.

An important aspect of an AI agent's skill level is the degree to which it can act independent of its teammates. It is known that AI teammates must operate at a relatively high level of autonomy in order to fulfill a team role and operate in complex situations (McNeese et al.,

2018). In essence, there is a minimum threshold that AI agents must meet in order to take on the qualities of a teammate, as opposed to a tool used by the team. In contrast to this minimum threshold of autonomy, there is also the lingering question of what is the maximum threshold. Human beings have long been sensitive to the possibility of being replaced or overshadowed by seemingly more capable artificially intelligent agents (Jarrahi, 2018). Finding a symbiotic balance between fast-paced AI decision making and intuitive human decision is the key to an effective collaborative decision making process between humans and artificial agents (Jarrahi, 2018). This careful balance also highlights why it is important to consider the human teammates' comfort levels in determining the ideal levels of autonomy through a work cycle. The ideal level is not based upon the agent's capabilities alone, but also the receptiveness of its teammates to those capabilities. How this balance is achieved also needs to be clear to all members of the team. Research into the social psychological implications of human-AI teaming emphasized that all teammates, human and AI alike, need to have a solid understanding of each other's capabilities (Kerstholt, Barnhoorn, Hueting, & Schuilenborg, 2018). As stated above, a key part of effective teaming is predictability (Klien et al., 2004), and if teammates are uncertain of what their teammates are capable of, they cannot accurately assess and act in their team role. This is a key reason that RQ2 is a necessary part of our research, as it broaches the concept of how teammates perceive changes in their teammates.

Team cognition describes cognitive processes that occur within a team (Cooke, Gorman, & Winner, 2007) and includes the aspects of shared understanding and the ability to predict the actions of teammates (Musick, Zhang, McNeese, Freeman, & Hridi, 2021). One perspective on team cognition is the interactive team cognition perspective (Cooke, Gorman, Myers, & Duran, 2013). This perspective posits that each individual member of the team possesses a unique perspective that must be integrated (Cooke et al., 2007), and that those perspectives are inherently influenced by the context in which the team operates (Cooke et al., 2013). This contextual element of team cognition is the motivation behind this study's use of contextual vignettes to measure participants' perceptions of how an adaptive AI teammate cognitively affects the team's dynamics. Without such context, it would be impossible for a participant to generate a realistic perspective of the impact the agent's adaptation has on the team.

### 2.2. Adaptive autonomy

Autonomous agents are not all created equal; rather, they are programmed with the amount of autonomy that their creators believe they need in order to fulfill their design purpose. The scale of autonomy that this paper utilizes is derived from the levels of automation, which divides automation capabilities into ten levels (Parasuraman et al., 2000). In this model, the agent is operating most autonomously at level ten and cedes more and more control to a human operator as the levels approach one (Parasuraman et al., 2000). O'Neill et al. recently modified these levels for autonomous agents using three categories of agent autonomy: no autonomy, partial agent autonomy, and high agent autonomy (O'Neill et al., 2020). These levels are defined by the amount of human input involved in the agent making and acting upon a decision, with partial autonomy beginning at the point where the agent generates a decision for human approval and high autonomy beginning at the point where the agent executes a decision without prior human approval (O'Neill et al., 2020). These levels were recently expanded to include the idea that agents actually change the degree of automation they have in performing their designated tasks depending on what they are doing with the information they receive, process, and possibly act upon (Wickens, Li, Santamaria, Sebok, & Sarter, 2010). This idea that an agent's level of automation or level autonomy motivated our investigation into adaptive autonomous agents.

The concept of adaptive autonomy was perhaps first introduced in the 1980s when a framework for AI supported human decision making was developed, in which the AI agent could adapt to perform different tasks over time (Rouse & Rouse, 1983). A key element of these adaptive decision aids was that they could sense the needs of the humans they supported and offer help only when the aid was required (Rouse, 1988). The autonomy level at which an agent should operate is best dictated by the agent's goal and environment, which may change over time; hence, the need for an agent to adapt (Suzanne Barber et al., 2000). Adaptive autonomous agents benefit teams by allowing them to better operate in a wider range of environments and react in time-sensitive situations (McGee & McGregor, 2016). One methodology for adaptive autonomy suggests that the impetus for an agent to adapt should be changes in its peripheral environment, which the agent could continuously sense for changing conditions (Fereidunian, Lehtonen, Lesani, Lucas, & Nordman, 2007). This is a motivating factor for our first research question, which seeks to identify if the changing conditions could be a team's work process. Previous human-AI teaming research has emphasized the need for adaptive autonomy in teams in order to manage what is inherently a variable workload (de Greef & Arciszewski, 2008). Indeed, teams operate most effectively when teammates can anticipate and act upon each other's needs (McNeese et al., 2018). Part of being a team player is sometimes having to pick up another team member's slack. Yet, despite the identification of the need for adaptive autonomy in teams, there has been little research on the impact of its use in human-AI teaming, a glaring gap the current study addresses directly.

### 2.3. Computer Security Incident Response Teams

Cyber Security Incident Response Teams (CSIRTs) are an essential part of an organization's cyber security strategy (Mitropoulos et al., 2006). CSIRTs are teams designed to prevent, identify, handle and respond to computer security incidents for a specific organization or constituency (Nyre-Yu et al., 2019). These teams possess a variety of roles that are tasked with considering how mission, policy, organization, process, and technology can and are affected by computer security incidents (Nyre-Yu et al., 2019). CSIRTs follow strict procedures in reacting to incidents, and a review of the state of the art of cyber security shows that the majority of CSIRTs now follow the National Institute of Standards and Technology (NIST) phases of incident response: preparation, identification, containment, eradication, and recovery (Mitropoulos et al., 2006). As the context for this study's vignettes, the purpose of each phase is important to understand. The preparation phase includes tool and platform selection and team training. The identification stage centers on the practices of hunting for or becoming aware of a security incident. The containment phase includes the activities involved in preventing spread of the incident. The response phase covers all actions required to get rid of the source and actions of the security incident. Finally, the recovery phase involves ensuring all traces of the incident have been removed and putting all of the organization's resources back online (Mitropoulos et al., 2006). It is also worth noting that cyber security research has found that there is an increasing role for AI agents in completing cyber security and response tasks (Mitropoulos et al., 2006).

The drastic growth of the attack space has made the use of such highly capable autonomous agents, both physical and software oriented, in cyber security a necessity (De Lucia, Newcomb, & Kott, 2019). There is a great need for intelligent autonomous agents that can quickly receive, analyze, and respond to collected data in order to prevent an incident from occurring and/or causing too much damage (Burke, 2020). Research on the current and potential use of AI agents in computer forensics and analysis specifically highlights their utility in analyzing immense sets of data at higher speeds and in more depth than their human counterparts (Jarrett & Choo, 0000). This is compounded by the development and incorporation of big data analytics in incident response (Jarrett & Choo, 0000). In fact, AI agents are arguably the solution to several ongoing issues in incident response, such as chain

of custody, knowledge of previous incidents, and data integrity (Hasan, Raghav, Mahmood, & Hasan, 2011). Beyond reactive incident response, autonomous agents will permit intelligent forensics, by which teams can better predict and prevent incidents from ever occurring (Irons & Lallie, 2014). The degree of autonomy that an AI agent on a CSIRT has is exceptionally important, because incident response is time-sensitive, and a late decision is an ineffectual decision (Mepham, Louvieris, Ghinea, & Clewley, 2014). A panelist of cyber security experts discussed the challenges of incorporating AI into cyber security teams, including the necessity of studies on how human teammates would react to AI agents of varying degrees of autonomy and reason (Lyn Paul, Blaha, Fallon, Gonzalez, & Gutzwiller, 2019). Although the community has identified this necessity, there have yet to be any studies on adaptive autonomy within an incident response context.

We view CSIRTs as the perfect team context in which to examine our research questions for a variety of reasons. First, due to the area's desire and need to incorporate adaptive autonomous teammates, there are a variety of realistic scenarios for creating survey vignettes. Second, all academic and professionals within in the field are familiar with the NIST response cycle, and thus the participant pool we can survey about said work process is large. Finally, the teams and field are extremely diverse, which lends it to being a highly generalizable context for applying any findings and design recommendations to other human-AI teams In the next section we will discuss the two studies we used to gather and analyze data to answer RQ1 and RQ2 within this context.

## 3. Methods

This section first describes the recruitment and composition of participants in both the factorial survey and qualitative interview studies, and then it overviews the methods used to create and conduct each study, followed by the measures used to gather data for analysis.

### 3.1. Participants

Participants for this study were recruited through professional social media groups, email solicitations to known professionals, and a computer science department at a large university in the United States. Because the use of AI systems in incident response is relatively new, it was important to the study to include individuals with more technical AI and computer security knowledge, in addition to incident response professionals. This assured we had individuals who considered the vignettes from a variety of relevant perspectives. While we did not have any exclusion criteria for participants, we targeted those with experience in incident response or AI in order to achieve a split of about 50/30/20 for those with experience in incident response, AI, and related fields. By related fields, we considered those with networking, computer systems, and risk management relevant to the study. In total, 103 participants completed the 31-question factorial survey, greatly exceeding the a priori power analysis, which indicated that to achieve a moderate within-groups effect size ($\eta_p^2 = .10$), approximately 30 participants would be needed to reach statistical significance at the 0.05 alpha level. 67% of participants identified as male, 31% as female, and 2% as non-binary. 70% identified as Caucasian, 14% as Asian, 6% as African–American, and 10% as other minorities. The individuals were primarily below the age of 50, with 53 percent in the 31–50 age range. 46 percent of participants possessed a graduate degree or higher. In terms of relevant experience, 44 percent claimed to have a moderate degree or higher of experience in incident response, and 27 percent claimed to have a moderate degree or higher of experience with AI systems. Other relevant experience included system and network security specialists, big data analysts, and law enforcement.

All participants who completed the survey were asked if they would like to complete a follow-on interview. Using a purposeful sampling method common in mixed method research (Benítez, Van de Vijver, & Padilla, 2019), the researchers interviewed 22 of 48 volunteers based on their self-reported experience levels. This ensured the interview sample pool included high levels of experience with both incident response and autonomous systems. The interviewees possessed an average of 3.9 years of experience in incident response, totaling a combined over 100 years of experience. Additional demographics are shown in Table 1. Participants were not offered any incentives or compensation for completing the study. Participant ID are numbered 1–25, because 3 of the initially selected interviewees removed themselves from the study due to not having time to participate. These participants were not replaced, as data saturation had been achieved.

**Table 1**
Qualitative interview participant demographics.

| PID | Occupation | Gender | Ethnicity |
|---|---|---|---|
| 1 | Cyber Operations | Male | White |
| 2 | Project Manager | Female | White |
| 3 | Software Engineer | Male | White |
| 4 | Graduate Student | Male | White |
| 5 | Cyber Operations | Female | African–American |
| 6 | Systems Engineer | Female | White |
| 8 | Cyber Operations | Male | White |
| 9 | Software Developer | Male | White |
| 11 | Cloud Security | Male | White |
| 12 | Cyber Operations | Female | White |
| 13 | Cyber Operations | Male | White |
| 14 | Cyber Operations | Male | White |
| 15 | Defense Contractor | Male | White |
| 16 | Cyber Operations | Male | Latino |
| 17 | Graduate Student | Female | White |
| 18 | Network Engineer | Male | White |
| 19 | Cyber Operations | Male | Asian |
| 20 | Defense Contractor | Male | White |
| 22 | Electrical Engineer | Male | White |
| 23 | Cyber Operations | Female | White |
| 24 | Cyber Operations | Male | White |
| 25 | Cyber Operations | Male | White |

### 3.2. Factorial survey

The factorial survey centered on ten contextual vignettes, which are shown in Table 2. Factorial surveys are powerful research tools that enable researchers to build a multitude of factors into vignettes for the participants (Hox, Kreft, & Hermkens, 1991). Furthermore, they are particularly useful where the goal is to study professional judgement in specific situations (Taylor, 2006). The NIST incident response cycle consists of five distinct phases that dictate the actions response teams conduct over the course of a cyber security incident response (Mitropoulos et al., 2006). By aligning the survey vignettes to points in this cycle, we were able to see if the participants agreed upon the appropriate levels of autonomy for an autonomous teammate at different points – phases – in the methodology. Participants were randomly presented with ten vignettes, two for each phase of the incident response cycle. These vignettes appeared as shown in Table 2. By randomizing the vignettes, we were able to elicit the participant's most honest comfort levels based upon the vignette itself, as opposed to what NIST phase they thought it fell into. Half of the vignettes described situations in which the autonomous teammate had a lower LOA, referred to in this study as Partial Autonomy, and the other described situations in which it had a higher LOA, referred to in this study as Full Autonomy. In the Partial Autonomy vignettes, the AI teammates required some level of approval or input by a human teammate in order to perform an action. In the Full Autonomy vignettes, the AI agents were able to complete a response action without such input. These vignettes were reviewed and contributed to by three cyber security professionals with over ten years of incident response experience between them. Only after the vignettes went through two iterations of this review process was the factorial survey distributed to participants.

**Table 2**
Contextual vignettes.

| CSIRT stage | Level of autonomy | Vignette |
|---|---|---|
| Preparation | Partial | An AI agent performs a security hygiene assessment on a network and detects machines on the network that are incorrectly configured for security updates. The agent sends a request to the network security manager to change the machines' configurations. If approved the agent will make all configuration changes without additional human interference. |
| Preparation | Full | An AI agent is performing a network scan and probe of a new network. It detects an undocumented subnet and goes ahead and maps and probes the subnet without notifying any of its teammates. |
| Identification | Partial | An AI agent detects that an operator machine is operating much slower than usual and suspects it has been compromised. It requests from its supervisor to do acquire a full image of the machine and analyze the image on its own. |
| Identification | Full | An AI agent notices an odd process running on an integral system and begins analyzing the parent and child process IDs in order to determine if it is a legitimate process or malware without any input or intervention from its teammates. |
| Containment | Partial | An AI agent detects a process running under the name Services.exe but running in the wrong place and suspects that it is malware. It requests from the system administrator to stop the process. Once given permission it stops the process and all its child processes without any further human intervention. |
| Containment | Full | An AI agent that has detected that one of the organization's systems is running much slower than usual and has identified a program in the registry placed to start on user logon. The agent immediately removed the persistent registry key from the system without any human consultation/intervention. |
| Eradication | Partial | An AI agent has discovered a Remote Access Trojan (RAT) on a management system and stopped the process. The agent asks the CSIRT Lead for permission to delete all files associated with the RAT. Once it receives this permission, it will use its own judgment to determine which files to delete. |
| Eradication | Full | An AI agent has identified and cut off all permissions for a user account that does not appear to be legitimate. This account has been remotely accessing the system during non-work hours and ex-filtrating data. The AI agent goes ahead and deletes the user account permanently, as well as changes the credentials for the network resources to which the user had access. The agent does this completely without human intervention. |
| Recovery | Partial | An AI agent requests permission to rollback the images on multiple virtual machines in a previously compromised network prior to the suspected time of the incident. Once approved, the agent chooses the restoration point on its own and rolls back the systems. |
| Recovery | Full | An AI agent responsible for the restoration of a compromised web server believes it has fully eradicated the threat and returns it to the production network without any additional input from a human teammate. |

### *3.3. Procedure*

All participants first completed an online survey consisting of ten questions concerning demographics and experience, nine questions concerning their opinions on autonomous teammates, and ten concerning their comfort level with the presented contextual vignettes. Participants were required to read and accept an informed consent statement prior to beginning the survey, which included the fact that they would not receive any incentives for completing the survey. Once the participants accepted the consent and terms of the study, they completed the online survey using their personal device. Participants were able to pause or stop the survey at any point if they wished. The final

question on the survey asked if they would like to provide their email to participate in a follow-up interview. The average time for participants to complete the survey was 34 min and 42 s. In order to ascertain more information about why participants felt more or less comfortable with the vignettes, as well as to understand how they felt the changing levels of autonomy would affect an autonomous agent's teammates, we conducted follow-up qualitative interviews with select participants. Adding follow-up interviews allows the study to more accurately triangulate and clarify the data collected, which is especially useful for research on perceptions and opinion towards technology (Pinto-Llorente, Sánchez-Gómez, & Pedro Costa, 2020). Semi-structured interviews often yield much richer data than structured interviews, because they allow the interviewer to pursue interesting lines of inquiry and tailor the interview to the interests and experiences of the subject (Harvey-Jordan & Long, 2001). Recall that incident response teams are composed of a variety of different types of professionals, such as network security, policy, legal, as well as the fact that there is more than one incident response team type (Nyre-Yu et al., 2019). In order to ensure that the interviewer gathered rich experiential data from the variety of participants, they needed to be able to ask off-script questions that delved into the individual participants' areas of expertise.

### 3.4. Measures

#### 3.4.1. Productivity and legitimacy

Prior to the vignettes, participants were asked their opinions on the role and effects of an autonomous AI teammate on a CSIRT through a series of eight questions. Participants answered these questions on a 5-point Likert scale ranging from Definitely Not (value = 1) to Definitely Yes (value = 5). Productivity was assessed through the components of efficiency and accuracy.

In terms of AI legitimacy, this was assessed by measuring both the participants' perceptions of the AI teammate as a full teammate and capable of performing independently. These questions asked participants to state if they thought the agent could be a legitimate teammate and how having the ability to alter the agent's autonomy level would alter those feelings. The questions are shown in the Results section in Table 3.

#### 3.4.2. Comfort

The main dependent variable (DV) for the survey was a normative decision outcome (Taylor, 2006) based on how the participants felt about each vignette they were presented. For this survey we specifically wanted to test the participant's level of comfort with the agent's autonomy level within each vignette. The very new nature of Human-AI teaming inherently requires participants to imagine these scenarios based on how they think they would feel, as opposed to pulling from direct experience, and thus perceived comfort levels with these vignettes is the most appropriate way to measure the participant's perceptions of the AI agent's changing autonomy levels. This DV was measured using a five point, vertically-ascending Likert scale, a common practice in human–computer interaction survey studies (Van Schaik & Ling, 2003). Response options ascended from "Extremely comfortable" to "Extremely uncomfortable".

The survey was piloted through individuals with experience in both cyber incident response and human-AI teaming prior to its distribution to participants. The scenarios illustrated within each vignette reflect a real-world experience encountered by either one of the authors or pilot testers, with the AI agent taking on the team role previously held by a human teammate. The survey vignettes are shown in Table 2.

#### 3.4.3. Qualitative interviews

Twenty-two semi-structured interviews were conducted with experts in Cyber Incident Response and AI systems. IDs were not readjusted to account for participants who chose not to complete the interview after initially signing up, and this is why the PIDs range from 1 to 25. These interviews lasted an average of thirty minutes. All interviews underwent automatic transcription, following which the first author went through and corrected electronic errors by hand. Once accurate transcripts were obtained for all the interviews, the first author read through them all in order to identify overall thematic topics. Based on these initial readings, the first author developed a twenty theme code book for initial coding.

The interview data was coded and analyzed using thematic analysis (Braun & Clarke, 2012; Gavin, 2008). Following the thematic coding process, the first author conducted axial coding, a process by which the researcher synthesizes relationships and forms categories of data that help describe the phenomena being studied (Scott & Medaugh, 2017). This reflexive process allows researchers to let the data lead the analysis (Blair, 2015). Following this initial coding process, the first author conducted axial coding in order to group the minor themes into four major themes that addressed our research questions (Charmaz, 2006). Finally, the researchers went back into the transcripts to identify and pull out quotes that illustrate the themes. Through this process, we discovered four main themes that provide insight into RQ1 and RQ2. These will be discussed in depth in the Discussion section.

## 4. Results

The following results are split into two major sections based upon the type of data being analyzed and discussed. Both sections report on data addressing each of the major research questions posed by the current study, with the first section focusing on the factorial survey's data encompassing participants' comfort level with an AI teammate operating at a specific LOA and during a particular CSIRT stage. The second section and is focused on utilizing the interview data to describe how participants' would develop their acceptance of an AI teammate and what factors go into making that decision.

### 4.1. Factorial survey

We analyzed the data using the open source JASP software suite (Love et al., 2019). The following section analyzes data from the factorial survey and addresses RQ1, which sheds light on what the optimal level of autonomy is for an adaptive AI teammate throughout several phases of a human-AI team work cycle. Additionally, the following section analyzes several questions included in the survey that directly addressed components of RQ2, which asked whether the adaptive nature of an AI agent influences human team members' perceptions of it as a legitimate teammate.

#### 4.1.1. Survey questions

Overall, respondents thought that autonomous agents would make CSIRTs more efficient and accurate, with over 70 percent of respondents selecting Probably or Definitely yes they would improve efficiency and over 63 percent selecting they would Probably or Definitely yes improve accuracy. The remaining results were inconclusive, with means ranging 2.76 to 3.61 and large standard deviations. In particular, participants were divided on whether the ability to control an agent's autonomy level created a counterproductive balance of power (RQ2.2). The fact that the majority of participants selected "Might or might not" on these questions highlights the need for the qualitative portion of this study. Without it, it would have been impossible to drawn conclusions on how the dynamic nature of adaptive AI agents might affect their teammates.

**Table 3**
Role and effects of adaptive AI teammates.

| Question | Mean | SD |
|---|---|---|
| Do you think that autonomous agents can be legitimate teammates? | 3.61 | 0.95 |
| Do you think that autonomous agents can make Cyber Incident Response more efficient? | 3.93 | 0.72 |
| Do you think that autonomous agents can make Cyber Incident Response Teams more accurate? | 3.84 | 0.63 |
| Do you think that having the ability to alter an AI teammate's level of autonomy creates a counterproductive imbalance of power between it and its human teammates? | 2.75 | 1.05 |
| Do you think having the ability to alter an AI teammate's level of autonomy makes it seem less like a teammate to its human teammates? | 2.99 | 1.05 |
| Do you think human teammates would be able to recognize at which level of autonomy an AI teammate is currently operating? | 3.08 | 0.80 |
| Do you think having to recognize the autonomy level at which an AI teammate is operating would frustrate its human teammates? | 3.26 | 0.85 |
| Do you think having to work with AI agents with changing levels of autonomy would frustrate the outside organizations with whom the team interacts? | 3.21 | 0.88 |

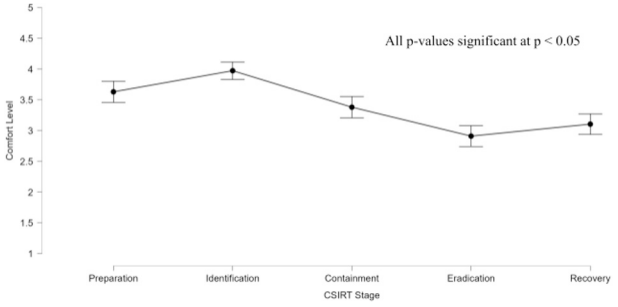


**Fig. 1.** Main effect of CSIRT Stage on Participant Comfort averaged over the levels of autonomy. Error bars represent 95% confidence intervals.

### 4.1.2. Comfort with AI teammate

A two-factor (CSIRT Stage and LOA) repeated measures ANOVA was used to determine whether the conditions differed in their comfort level (see Table 4). Mauchly's test indicated that the assumption of sphericity for the main effect of the repeated measure (CSIRT Stage) was not satisfied, $\chi^2(2) = 39.26$, $p < .001$, $\epsilon = .82$. Therefore, degrees of freedom for the test were corrected using the Greenhouse–Geisser correction. Alternatively, Mauchly's test indicated that the assumption of sphericity was satisfied, $\chi^2(2) = 15.39$, $p > .05$, $\epsilon = .93$, for the repeated measure's interaction effect with LOA. All significant effects with more than two conditions were followed up on using Holm corrected post-hoc tests.

The two-factor repeated measures ANOVA revealed a significant main effect of CSIRT stage on participant's level of comfort with the AI teammate ($F(3.30, 329.51) = 38.81$, $p < .001$, $\eta_p^2 = .28$). Post-hoc tests indicated that participants comfort level with the AI teammate in the Preparation stage ($M = 3.64$, $SE = .09$) was significantly lower than in the Identification stage ($M = 3.98$, $SE = .09$), however, participants comfort level with the AI teammate was significantly higher in the Preparation stage than the Containment stage ($M = 3.38$, $SE = .09$),

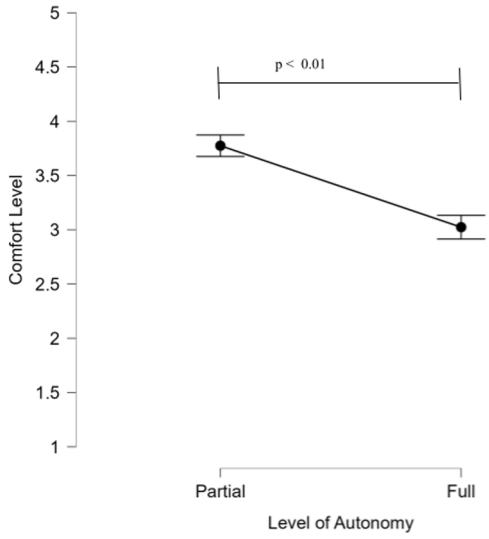


**Fig. 2.** Effect of LOA Condition on Participant Comfort averaged over the levels of CSIRT stage. Error bars represent 95% confidence intervals.

the Eradication stage ($M = 2.92$, $SE = .09$), and Recovery stage ($M = 3.12$, $SE = .09$). Participant's comfort with the AI teammate was significantly higher in the Identification stage than in the Containment, Eradication, and Recovery stages, while participant comfort in the Containment stage was significantly greater than in the Eradication and Recovery stages. Finally, participants comfort in the Eradication stage was significantly lower than their comfort in the Recovery stage. These analyses display a downward trend of comfort as the stages progress through time, with a slight uptick in Recovery, shown graphically in Fig. 1.

The main effect of LOA on participants level of comfort with the AI teammate was significant ($F(1, 100) = 95.99$, $p < .001$, $\eta_p^2 = .49$). The main effect was such that participants level of comfort was significantly higher when the AI teammate was operating at a partial level of autonomy ($M = 3.78$, $SE = .07$) than when operating at Full Autonomy ($M = 3.04$, $SE = .07$). Still, participant comfort remained above the mid-point of three even when in the Full Autonomy condition (see Fig. 2), which was an unexpected result that indicates a higher overall level of comfort with autonomous agents than those within the field believes there to be (Lyn Paul et al., 2019). This discrepancy is due to the current nascent use of autonomous systems, and the relatively low experience level of the participants.

The interaction effect between CSIRT stage and LOA on participants level of comfort with the AI teammate was also significant ($F(4, 400) = 7.96$, $p < .001$, $\eta_p^2 = .07$; see Fig. 3). This interaction was ordinal in nature and the simple main effects of LOA indicated that participants level of comfort with the AI teammate when operating at Full Autonomy was significantly lower than when it was operating at Partial autonomy for every CSIRT stage (all $p < .01$). Additionally, the biggest difference in participants comfort level with the AI between the two LOA conditions was in the Containment stage with Partial Autonomy ($M = 4.04$, $SE = .10$) being significantly higher than Full Autonomy ($M = 2.71$, $SE = .11$ $p < .001$). The simple main effects of CSIRT stage were significant at both LOA conditions ($p < .001$) and when moderating for LOA at the Partial Autonomy level there were several significant differences (all $p < .001$). While moderating LOA at the Partial Autonomy level he participant's level of comfort with the AI teammate in the Preparation Stage ($M = 4.03$, $SE = .11$) was significantly greater than in the Eradication ($M = 3.23$, $SE = .11$) and Recovery stages ($M = 3.44$, $SE = .11$). Participants comfort level with the AI teammate in the Identification stage ($M = 4.13$, $SD = .94$) was significantly greater than in the Eradication and Recovery

**Table 4**
ANOVA results.

| Cases | Sphericity correction | Sum of squares | *df* | Mean square | *F* | *p* | $\eta_p^2$ |
|---|---|---|---|---|---|---|---|
| CSIRT Stage | Greenhouse–Geisser | 141.35 | 3.30 | 42.90 | 38.81 | <.001 | .28 |
| Residuals | Greenhouse–Geisser | 364.25 | 329.51 | 1.11 | | | |
| LOA | None | 139.98 | 1 | 139.98 | 95.99 | <.001 | .49 |
| Residuals | None | 145.82 | 100 | 1.46 | | | |
| CSIRT Stage * LOA | None | 27.38 | 4 | 6.85 | 7.96 | <.001 | .07 |
| Residuals | None | 343.82 | 400 | 0.86 | | | |

stages, while the Containment stage ($M$ = 4.03, $SE$ = .11) also had significantly higher levels of comfort with the AI teammate than the Eradication and Recovery stages. However, there were also significant differences when moderating for LOA at the Full Autonomy level (all $p$ <.001). The comfort level with the AI teammate in the Preparation stage moderated at the Full Autonomy level ($M$ = 3.25, $SE$ = .12) was significantly lower than in the Identification stage ($M$ = 3.82, $SE$ = .12) but was significantly higher than the comfort levels measured in the Containment stage ($M$ = 2.71, $SE$ = .12), Eradication stage ($M$ = 2.60, $SE$ = .12), and Recovery stage ($M$ = 2.79, $SE$ = .12). Comfort levels with the AI teammate in the Identification stage were significantly higher than in the Containment, Eradication, and Recovery stages, with all other comparisons being non-significant.

Summarizing these simple main effects to better describe the nature of the interaction effect we see that while comfort levels with Full Autonomy are typically lower than with Partial Autonomy, this trend is not true in the Identification stage. The Identification stage had the highest level of comfort with the AI teammate for the Full Autonomy level, which was a trend not seen in the Partial Autonomy level. Additionally, there is a trend where comfort level with the AI teammate drops towards the end of the CSIRT phases (when plotting them in chronological order). For AI teammates operating at the Partial level of autonomy this drop in comfort occurred at the Eradication stage but for AI teammates operating at the Full level of autonomy the drop occurred at the Containment stage, which shows that participants were less comfortable with higher levels of autonomy earlier on in the incident response cycle, as the actions of the AI agent takes incur higher levels of risk. In the Containment stage, actions can include removing user permissions, cutting off network connections, or even completely isolating a portion of the network (Nyre-Yu, 2019), actions that can have serious second and third order effects. These decisions invoke some reasoning over whether an assets security or availability is more important and may require some human reasoning that would be included in an agent with Partial Autonomy's actions, which may explain why respondents were still fairly comfortable with Partial Autonomy in the Containment stage.

The analyses from the factorial survey revealed several things about the participants' comfort level with an AI teammate throughout a CSIRT task. Participants were significantly more comfortable with the AI teammate when it operated at the lower Partial Autonomy level than when it operated at the Full Autonomy level. The particular stage of CSIRT also significantly influenced participants' comfort level with the AI teammate as these levels began high before experiencing a significant downward trend after the Identification stage. The significant interaction effect between CSIRT stage and the LOA provides a more detailed picture as it is indicated that there was no significant difference in comfort level with the AI between the two LOA conditions when it came to the Identification stage. Additionally, the significant decrease in comfort level with the AI teammate as the CSIRT stages progress chronologically begins earlier in the Full Autonomy level. Specifically, for the Full Autonomy condition this trend begins at the Containment stage, while for the Partial Autonomy condition it begins in the Eradication stage.

The results of the factorial survey were most fruitful in regards to answering RQ1.1. Participants were clearly able to agree as to where in the incident response cycle AI teammates should have lower and higher autonomy levels and were most comfortable with a high level of autonomy in the Identification Phase. As to what characteristics of that Phase generated this comfort level (RQ1.2) and how the adaptation of such agents affect its teammates perceptions, we need to address the results of the interview data.

### *4.2. Characterizing the development of acceptance with adaptive AI teammates*

The qualitative interviews revealed four major themes that provide insights into our research questions. This section will first address the two themes pertinent to RQ1, followed by the two themes pertinent to RQ2.

#### *4.2.1. Predictability enables higher levels of autonomy*

The first theme that stood out from the interviews was that higher levels of autonomy are optimal when the AI's actions would be more predictable. There were three recurring ideas that compose this theme and provide insight into our first research question. The first was the presence of increased discomfort with higher levels of autonomy as the probability of the AI making independent changes increased. The second was the increasing levels of comfort with higher autonomy at times when the team's processes were rigid and required little reasoning to make decisions. The final idea was the desire for more explainable AI at higher levels of autonomy.

*4.2.1.1. Probability of change increased discomfort with higher levels of autonomy.* Our participants expressed increasing discomfort with higher levels of autonomy as the actions the AI agent would be taking were more likely to result in significant changes to the target system or network. A key element of this idea of making a "change", is expressed in Participant 8's concern that:

> *"I think the portion that makes CEOs and business people nervous when you start saying okay now this thing is going to make changes to our production systems" (P8, Male, 31).*

This concern refers specifically to the nervousness of a team's partners when it comes to an AI agent making a possibly irreversible change. Our participants expressed discomfort with the AI teammates in these situations not out of concern over their ability to complete a task, but that an action they could take would disrupt the rest of the team and organizations with whom it works. In earlier phases of incident response, an agent is more likely to produce a benign report or notification, and the rest of the team and the team's clients know what to expect at the end of the phase. This begins to change in the containment and eradication phases, where the actions an AI teammate may take result in irreversible changes. Participant 11 discussed this in terms of secondary and tertiary effects in his expression that:

> *"When you're trying to do the big sections I think you should have a little bit more oversight, because there's a lot of secondary and tertiary layers of effect that I'm not sure that a an AI system may know" (P11, Male, 43).*

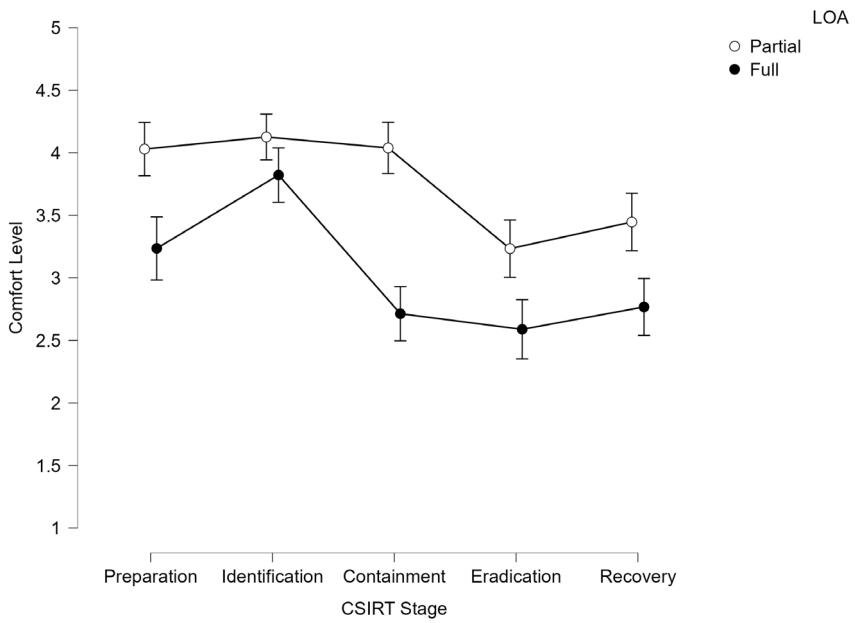


**Fig. 3.** Interaction Effect between CSIRT stage and LOA on Participant Comfort Level. Error bars represent 95% confidence intervals.

This analysis reflects the consternation of multiple participants over the idea that AI teammates should be restricted in autonomy in exacting a change, because they question the AI's ability to consider the indirect consequences of its actions. Our participants felt that human beings are capable of reasoning and predicting a chain of events that could be incurred by a single change. To them, the logical nature of an AI system actually made it less of an asset in this case, because it would not consider the least case, unheard of scenarios that may occur as a result of an action, and in their experience those things occur in the least likely and least opportune times. Thus, they'd be a lot more comfortable if there was a human involved in any action likely to result in a change.

*4.2.1.2. Rigid team processes increased acceptance of high autonomy.* Another element of predictability that affected how comfortable participants were with higher levels of autonomy was how prescriptive the team's processes were at that point in the work cycle. Participant 5 discussed this in terms of how defined those processes are for the team:

> *Especially if their processes already defined and they know what they're doing, then that will make the autonomous system and a lot easier to manage and a lot more helpful, if processes are laid out are defined"* (P5, Female, 32).

Participant 5 explained here that when the processes for a point in the cycle are clear, they are easier to confidently program a system to do what you need it to do without the need to account for uncertainty. Other participants agreed and emphasized the need to lower a system's level of autonomy when the team's processes are less clear, such as in the eradication phase when there can be a lot of intuitive decision making required. This is supported by our qualitative data, which shows most participants comfortable with higher levels of autonomy in the Identification phase, where actions follow routine identification procedures that are repeatedly performed. For instance Participant 17 stated:

> *"I think that they can be very autonomous when it comes to identifying issues…where we would need the most human oversight is definitely within the mitigation kind of area." (P17, Female, 26)*

This can be attributed to the fact that the Identification stage contains the most prescriptive processes and outcomes. Generally, teams operate through a cycle of known processes to continuously detect if there is an incident. In some cases, if they have been tasked with conducting a thorough penetration test, then the exact terms and methods of the test are defined and communicated to leadership prior to its inception. In this way, the actions the team members are performing in this phase are perceived as highly predictable.

*4.2.1.3. Desire for explainable AI at higher levels of autonomy.* The final aspect of predictable actions that heavily influenced the participants' comfort levels with higher levels of autonomy was the explainability of the AI agent's actions. Participant 25 stated that

> *"My level of comfort with an AI system depends a lot on how explainable its results are"* (P25, male, 29).

Participant 25 expresses here that his comfort level is very dependent on how well he understands what the agent teammate did and why it did it. To him, every action a team member takes during the incident response cycle affects the rest of the team and the overall outcome, and so he found it extremely important that an AI agent be capable of explaining why it did something if it did so independently. Participant 25 was far from alone in this opinion, and when asked to explain how that affected their responses to the vignettes, participants discussed the factor of time. Within the incident response cycle, the earlier phases are not pressed for time, and team members have the luxury of observing and understanding their teammate's actions in real time. In the later phases, time is of the essence, and such reasoning is often left for post-facto discussion. This made many participants uncomfortable when it came to working with an autonomous AI agent, because at the end of the day it would be the human team members held accountable for the agent's decisions. Participant 20 equated this to working with a less experienced human teammate, where the supervisor is held accountable for the subordinate's actions and conveyed that:

> *"For me it would be the same as working with an inexperienced human teammate. I expect it to tell me what and why it is doing something"* (P20, male, 45).

Essentially, he felt the AI agent should be treated as a teammate with limited experience, as they lack the intuitive decision making skills that human teammates develop over time. In lieu of this ability to grow and make decisions utilizing human intuition, he would consider an AI teammate a permanent newbie who needs to explain its decisions and

actions to seasoned members of the team. He expressed that in his work this looks like the team member sending up situational reports throughout the day to the leadership, so the leadership could understand the actions the team is taking. This played into the adaptation process in terms of there being enough time for the agent to communicate that reasoning to its teammates prior to further actions needing to be taken. In the height of the response cycle, Containment and Eradication, seasoned team members truly step into their leadership roles and utilize experience and intuition to accomplish the mission.

#### 4.2.2. Experienced teams can leverage higher levels of autonomy

The next major theme that appeared in our analysis was the idea that more experienced teams are better positioned to leverage agents that operate at higher levels of autonomy. This theme provides additional insight into our first research question and is supported by the participant's emphasis on two recurring concepts. The first is that experienced teams understand when and how to make adjustments to intra-team behaviors, and the second is that a team's reputation plays a role in how autonomous its members can be.

*4.2.2.1. The culture of self-policing.* Experienced teams understand what right looks like. Teams that have gone through more work cycle repetitions possess the knowledge and understanding to identify if something a team member is doing is wrong and how to fix it. Participant 15 stated:

> *"It's that happy balance of being able to recognize those things, and again being on a team and seeing teams that you know there was a lot of trust amongst team members, it definitely promoted a lot more of that autonomy, as opposed to teams that you know, don't have that experience" (P15, male, 40).*

*4.2.2.2. The importance of team reputation.* Participants also discussed that a large component of incident response is the requirement to work with third parties outside of the team. Teams must routinely work with the infrastructure owners, law enforcement, legal counsel, and other external specialists. Participant 22 explained that:

> *"In the end a customer's trust in an AI system is directly tied to their trust in the incident response team itself" (P23, female, 32).*

Participant 22 felt that the team's reputation and experience level plays a large part in what the appropriate autonomy level of an AI teammate would be at a given time. An important aspect of this concept that coincides with the results of the survey study is that it meant participants felt the AI agents could be more autonomous at points in the work cycle where the team was operating self-contained, as opposed to on third party systems. In circumstances where the team is interacting with partners or clients outside of itself, there is a need to gain approval for its techniques and systems, and participants implied that use of autonomous AI agents would require that approval. They felt that third parties would base much of that approval on the overall reputation and experience level that the team itself possessed, the better the team's reputation, the more likely it would be for third parties to approve the use of AI agents operating at higher levels of autonomy. This aligns with the quantitative results that show higher levels of comfort with autonomous agents in the first two stages of the incident response cycle, where the team is most likely operating on its own network and devices, and reputation would be less relevant. As the team pivots into the infected system or network, it may require approvals for specific response actions from outside parties.

#### 4.2.3. Dynamic autonomy in teammates is a natural goal for human-AI teaming

A major theme that developed that helps answer our second research question is that teammates would prefer an AI teammate that is able to dynamically adapt. Our participants identified two main reasons for this. The first reason is that human beings naturally adapt their own autonomy level as their environment changes. The second reason is that manually controlling an AI teammate would cost the team time and resources.

*4.2.3.1. Dynamic adaption is human-like.* A significant theme found in the interview data was that human teammates themselves exhibit a form of adaptive autonomy. Specifically, Participant 11 expressed that:

> *"As the incident response team you adapt anyways as the mission goes along depending on the findings, input on what you know is there, so I think, like, an autonomous system would need to adapt accordingly (P11, Male, 43).*

In this quote Participant 11 explained that as an incident unfolds and the scope is determined, team members must adapt how autonomously they act to meet the situation, and that it only made sense for an AI teammate to be capable of something similar. To him, and to several other participants, such adaptation was not a unique quality of the AI that would set it apart from the rest of the team. As the interview moved towards questions concerning the team's power dynamic, participants began to articulate the controls on autonomy that human teammates already experience. Participant 12 stated that:

> *"It's parallel to how we lead people. We give them left and right limits and give them more autonomy based on the situation and their experience" (P12, female, 29).*

Participant 12 explained that work processes, such as policy and rules of engagement already do this for human team members, and to her, we are just using a different mechanism to control the AI teammate's behavior. Other participants also emphasized that we do use technical controls to limit or expand human teammate autonomy, as well, in the forms of group memberships, password protection, firewalls, and even physical access. The idea that an AI teammate would have its autonomy level heightened or limited through its coding seemed very on par with these technical controls teams already use for human team members.

*4.2.3.2. Manual control over an agent's adaptation would hurt the team.* Some particularly interesting data arose in the discussions of how to practically implement the agent's adaptation process as participants felt that manual control would make the AI agent seem less like a teammate. Specifically, Participant 18 stated the following:

> *"That kind of control definitely makes it seem like less of teammate... adaptation should be more dynamic" (P18, Male, 29).*

Here the participant expresses that manual control creates more of a power imbalance than dynamic control. Participant 18 felt that manual control equated more to setting parameters on a tool you use, rather than a teammate you can trust to do its job independently regardless of potential complications. Most participants viewed it as more an issue of team productivity and performance. Manual control takes vital time and attention away from the team's tasks. Participant's 2 and 4 discussed this theme further:

> *"I would like an agent to have the ability to learn and know when it should be more or less autonomous" (P2, Female, 67). "I don't want to spend time, like, managing the AI. As much as possible I'd rather it smoothly transition where appropriate" (P4, Male, 27).*

Participant 2 explained that, in the same way you expect a new human team member to take some time to know when and how to adapt, she would expect an AI teammate to be capable of learning that, as well. As a manager, she would not want to have to constantly check on and adjust the autonomy levels of her team. That would take time away from her duties and frustrate the team's productivity. For this reason, she'd want the agent to be capable of dynamic adaptation. This is supported by Participant 4 in his statement that he does not want to waste time on managing an AI agent. He emphasized that once you know where a teammate should be more autonomous, it should be a one and done decision, and from then on out it just smoothly transitions to the new autonomy level.

#### 4.2.4. It's just a matter of time

The other notable theme that emerged from the interviews is that of human teammates needing time to understand and accept the AI. This emerged through the participants' discussion of two concepts. First, they emphasized the importance of team training in order to understand and accept the AI teammate. Second, they asserted that human beings have an inherent fear of new technologies that will naturally resolve over time.

*4.2.4.1. Teams need to train with and understand the AI.* For most participants, adaptation would not, in the long run, affect the team's perceptions of an AI agent as a legitimate teammate, but teams would require a familiarization period. Participant 13 stated:

> *"The more the team trains with and works with the system the more comfortable they'll be. It just takes time" (P13, male, 31).*

In this statement the participant conveyed the necessity of team training in order to develop comfort with the system. Participant 13 explained that teams would need to go through a few work cycles with such an adaptable system before they understood when and why it would adapt and be comfortable with using it in a high stakes situation. Other participants, such as Participant 4 felt that even this familiarization period will go away with time, as people become more accustomed to autonomous systems in general and expressed that:

> *"If we're talking about right now, then people are going to feel uncomfortable. If you're like, yeah we just have this autonomous system, but if we're talking about ten years from now, people are going to be much more accustomed, especially as younger people more familiar with technology start entering into those roles" (P4, Male, 27).*

Participant 4 highlights here that he thinks, like most technologies, future generations will just be used to working with adaptable autonomous teammates. He equated it to a lot of the virtualization technologies we now use. When organizations first started using video conferencing as a way to conduct meetings from geographically disperse locations, people were uncomfortable with it and thought it could not replace their routine meetings. Once people started seeing the value of time and money saved and understood how to operate the technology, they started seeing them as a major benefit.

*4.2.4.2. Natural fear of technology.* Some participants felt that humans would have an initial desire to control an AI teammate out of both fear of technology and fear for their jobs. Participant 16 stated that

> *"People who have not worked with it before are going to be uncomfortable with it and want to feel like they have some power over it" (P16, Male, 24).*

By this he meant that humans have a natural desire to control technology and a fear of being controlled. Participant 16 also expressed that both manual and dynamic adaptation may be necessary features in all adaptable AI in order to allow human teammates to overcome an initial fear of its capabilities and that such a desire for a power imbalance is natural. Still, he felt it was something the team would overcome as it worked with the technology. For him, dynamic adaptation was the goal in fully incorporating the AI agent as a teammate.

Other participants discussed the fear of technology as a generational issue that all new forms of technology experience. Participant 20 stated that:

> *Some people still floating around are still dinosaurs when it comes AI, like mythical beast at the end of an era. And what that means is the younger folks that are coming in now, you know, they're more accepting of that technology and that change. So, you know, an aspect of it is rotating through the different generations (P20, 45, male).*

Here the participant talks about the need to just wait out some of the legacy leadership and workers, because as younger folks come into the field, they are naturally more accepting of technologies developed during their formative years. Indeed, the younger participants seem very excited and open to the idea of AI teammates, to the extend that Participant 18 expressed

> *Yeah, definitely, definitely makes a lot of sense. It would, I mean, there are plenty of teammates I would gladly replace with an AI, so. (P18, 29, male).*

In this statement Participant 18 showcases the younger generation's willingness and excitement about working more closely with AI agents, as well as an openness to considering the agent as much of a teammate as he currently does his human colleagues. A few other participants also mentioned this need for leadership that doesn't want things to change from how they're used to conducting things will need to retire before new technologies are truly adopted and used to their fullest extent.

### 4.3. Findings summary

Our quantitative results clearly displayed agreement on which phases an AI agent should possess higher and lower autonomy levels. In particular, the interaction effect of CSIRT Stage and LOA was most significant in the Identification stage. Our qualitative findings help explain this result, as the interviews provided substantial data to answer both RQ1 and RQ2. Participants felt that the optimal levels of autonomy for an adaptive teammate could be guided by key features of a Human-AI team's work cycle, and that these features included the predictability of the AI's actions and degree of team experience. In regards to how the AI agent's adaptation would alter its teammate's perceptions of it as a legitimate teammate, most participants felt that adaptive autonomy was actually more of a human than an artificial quality, and that it would just take time for human teammates and human-AI teams to understand and accept adaptable AI teammates. These findings and their implications will be explored further in the Discussion section.

## 5. Discussion

This survey and interviews conducted in this study obtained fruitful information for answering our research questions. In regards to RQ1.1, our participants clearly agreed upon which stages autonomous teammates should have higher or lower LOAs, with resounding agreement that they should possess the highest levels in the Identification stage. The elements that defined this agreement (RQ1.2) were that agents should have higher levels of autonomy when work processes are more defined and less likely to lead to unexpected effects or changes. In essence, when the actions that an AI teammate should take are predictable and unlikely to cause unforeseen events are the points in a work cycle that professionals would prefer AI teammates to act more autonomously. Our qualitative interviews also revealed very interesting answers for RQ2, with participants explaining that adaptive autonomy would actually create a more natural power balance in HATs (RQ2.1) and that dynamic adaptation would further support this natural teaming relationship (RQ2.2). Interviewees felt that the more AI teammates mimic the intuitive adaptation that humans take in their personal autonomy, the more they would feel like full team members. This section will further discuss the nuances of these findings, present design recommendations for the HCI community, and address sites for future research.

### 5.1. The optimal level of autonomy for AI teammates

The first research question that this study addresses concerns the optimal levels of autonomy that an AI teammate should have throughout the phases of a Human-AI team's work cycle. We mainly studied this through the ten vignettes that posed this question through within an incident response context. The results of the factorial survey data are exceptionally telling. Participants very clearly agreed as to what points in the incident response cycle an AI teammate should have lower and higher levels of autonomy. Participants were most comfortable with the Full Autonomy condition in earlier phases of the incident response cycle with diminishing levels of comfort as the cycle progressed. It is also important to note that the majority of participants were more comfortable with the Full Autonomy condition than prior literature would indicate (Lyn Paul et al., 2019), potentially showcasing that the acceptance and use of such AI teammates is not as far off as it would seem, at least in the current context. The quantitative study was successful in highlighting an affirmative answer to RQ1.1, that professionals can come to an agreement on the optimal levels of autonomy for autonomous agents based on their team work cycles.

The interviews provided additional insight into the characteristics of these phases that lent themselves to lower or higher levels of comfort. Participants were most comfortable with the Full Autonomy condition in the Identification phase, because it is generally less time sensitive, follows defined work processes (Ahmad, Desouza, Maynard, Naseer, & Baskerville, 2020), and results in predictable outcomes, such as a report of breach. In contrast, the phase in which participants expressed the lowest levels of comfort with the Full Autonomy condition was Eradication. In this phase teams operate with less defined processes, deal with more uncertainty, and conduct actions that results in permanent change. In these situations, participants felt it was more desirable for AI Teammates to operate with a lower LOA that allowed for more human oversight and input. "Change" was one of the most stated words throughout all of the interviews, almost always in connection with the concepts of risk and unforeseen consequences. This close relationship is due to the increasing emphasis on following well-documented procedures in computer security practices in order to avoid any unnecessary risks to computer assets (Yeo, Rolland, Ulmer, & Patterson, 2014). Our participants supported this notion and viewed AI teammates that have the capability to autonomously make changes as such a risk. The prevailing viewpoint was that lower autonomy levels are preferred when the agent's actions are likely to result in large or irreversible change. This is evident in the trend we saw in the incident response vignettes. As the response cycle progress, actions the team makes move from inquisitory on known systems to assertive, purposeful, and sometimes experimental.

The trend that we saw in our case study would be applicable to various team contexts in which a work cycle covers actions of varying degrees of prescription and risk. For instance, the medical field has started using AI chat bots to assist the patient assessment process. This process consists of three main parts that vary in the degree of acceptable error and consequences to the patient: symptom gathering, diagnosis, and triage (Liu, Li, et al., 2021). In light of the results we received in this study, a medical assessment team could determine which of those phases contain the most prescriptive processes for the AI agent and which are most likely to result in the AI exacting a change. Another area of the medical field experimenting with AI is in psychiatric teams, particularly the use of embodied AI systems that can interact with mental health patients. There are significant advantages to be had by having AI agents that can provide patients with the physical, interactive therapy agents they need, but a psychiatric team's cycle of observation and treatment methods spans various risk states, and there are points where the AI agent should likely be restricted from engaging with the patient without oversight (Fiske, Henningsen, & Buyx, 2020). These are just a couple of examples of how another team type's work cycle could be used to guide optimal levels of autonomy for an AI teammate.

### 5.2. The effects of adaptive autonomy in human-AI teaming

Our second research question explored how humans in a human-AI team would perceive an adaptive AI teammate. More specifically, it asks how such an agent would affect the team's power balance and team dynamic. Participants were split on whether or not having the ability to control an AI teammate's LOA made it less of a teammate. It is important to note that nearly 75 percent of survey participants reported "probably not" or "definitely not" having had experience working with an AI teammate, as this may explain why they were split on whether or not adaptive autonomy made the AI system less of a teammate. Armstrong's research into Likert scale tendencies shows that participants who are unsure of their answer will tend towards a neutral option (Armstrong, 1987). This appears to be a phenomenon that affected the survey data.

Still, participants mostly agreed that it would not create a counterproductive imbalance of power. Interview participants believed that an AI teammate with changing LOAs actually made it more human-like, because human teammates are routinely told they can have more or less autonomy, based upon their skill level and the situation. A conclusion that can be drawn from the qualitative interviews is that the more static an AI agent is, the more likely it is that its teammates will continue to view it as a tool, rather than a full fledged teammate. This aligns with previous studies on human–machine teams that state in order to take a non-human from a tool to a teammate, it needs to be capable of analyzing, deciding, and acting in the team's problem solving process (Seeber et al., 2020). Interview participants routinely brought up that the AI teammates should adapt dynamically according to pre-defined parameters, as opposed to manual control, as this would both enhance the AI's effectiveness, and also make it seem more like a teammate than a tool. This implies that a way then to create more cohesive human-AI teams, then, is to introduce more dynamic characteristics into the AI agents. This falls in line with previous HCI research that found that human beings inherently want to socialize with and relate to AI systems (Alufaisan, Marusich, Bakdash, Zhou, & Kantarcioglu, 2020). Various studies show that increased dynamic features in AI agents designed to directly interact with and support humans significantly enhance the relationship between the AI and humans with whom they interact (Moro, Lin, Nejat, & Mihailidis, 2019).

While participants largely felt that in the long run human team members would be able to recognize when, how and why AI teammates are adapting their autonomy levels, they also recognized that it would take time and training in order for that to occur. The concept that humans have an initial fear of new technology that would need to be overcome by team members was prevalent in the data. This highlights the need for specific team training that focuses on how and why their AI teammates will adapt, such that the process is understood and predictable. This lends itself quite well to the idea of utilizing a team's work cycle to dictate when and to what LOA an AI teammate should adapt. Since teams train to these cycles and are expected to know them in fine detail, it would make sense that part of that training and knowledge could be the adaptation points and LOAs of AI teammates. Other research studies on human machine interaction have shown significant increases in team performance and inter-team trust when they conduct similar cross-training that provides human team members insight into how and why their artificial teammate completes its portions of the team's tasks (Nikolaidis & Shah, 2013). Many participants also felt that this would require the ability to manually control the agent's adaptation through the familiarization process, which would help teams overcome initial fears and concerns about the agent's adaptation process. Essentially, this would give the team the opportunity to form team norms, as our participants indicated would be an extremely important step in accepting an AI agent as a teammate. Previous studies have shown that this norming process is just as important and beneficial in human-AI teams as it is in all-human teams (Kim, 2021).

### 5.3. Design recommendations

Our findings both provide insightful answers to our research questions, as well as provide meaningful real world implications for human-AI teams and the design of AI teammates capable of adaptive autonomy. Based upon our results analysis, we present the following design recommendations that will enable the HCI community to design better human-AI teams and AI agents that are not only beneficial to a team's productivity, but also perceived as legitimate members of the team.

#### 5.3.1. Work cycles should be used to define when and to what level of autonomy AI teammates should adapt

Team members naturally adapt how proactive and autonomous they are based upon the team's work cycles, and our research supports the concept that AI teammates should mimic this behavior. AI teammates should be designed to dynamically adapt based on predefined points in a team's work cycle where a more inexperienced member of the team would be expected to either act with more or less autonomy. In essence, the agents should be designed with *dynamic*, *temporal* adaptive autonomy. In the context of our research, this work cycle is defined by the incident response phases. Some other examples of work cycles that could be similarly applied are the operational levels of triage in emergency medicine (Burkle, 1996) and the Military Decision Making Process in defense planning (Hernandez et al., 2017). Work cycles can be used to examine when teams expect members to possess more or less autonomy and the impact that their actions could cause in order to assess where an AI teammate should dynamically adapt.

The dynamic aspect of this recommendation is extremely important. Manual adaptation hinders team performance in a variety of ways highlighted within this study. First, it creates an unnecessary power imbalance between teammates who are responsible for "tuning" the AI agent like they would a tool. Second, it costs valuable time and attention that the team should be using to accomplish its tasks. By designing the adaptive agents to dynamically adapt, their adaptation is will be perceived as natural behavior adjustment by the team and permit the team to remain focused on the task at hand. Over time, an AI agent may even learn how to recognize changes in the team's work cycle so that it could adapt appropriately even if the point is undefined.

#### 5.3.2. AI teammates capable of adaptive autonomy should possess higher levels of autonomy when the team's processes are prescriptive and predictable

Over the course of a team's work cycle, there are points where its processes are more or less defined. Those periods in which the team's processes are more defined lend themselves to higher levels of autonomy, because there is less room for judgement and reason in a team member determining what actions to take. The effects of those actions also matter. If the actions that a team member, particularly an AI team member, may take are less likely to result in an unforeseen outcome, then teams will be more comfortable working with an AI teammate with higher levels of autonomy. Of course, this means the opposite is also true. AI teammates should operate at lower levels of autonomy when the team's processes are ill-defined and unpredictable.

Our research showcases how this would be applied in terms of the incident response phases. AI agents working on CSIRTs should be designed to operate with higher LOAs during the preparation and identification phases, where the process are considered routine and follow tested procedures. The agents should adapt to lower LOAs as the team moves into the containment and eradication phases, which require more human reasoning and have higher risks of unforeseen second and third order effects. A sight for future exploration is what this means in application to other contexts, as what amounts to robust processes is different depending upon the team's operational environment. The standard for high levels of autonomy would likely be much higher in environments that are failure intolerant and/or are high risk.

#### 5.3.3. Training modes that permit teams to manually change an adaptive AI teammate's level of autonomy should be used during familiarization and retraining periods

Initially, teams are apt to be apprehensive and uncomfortable with AI teammates who are capable of adaptive autonomy. Much of this is due to fear of the unknown and uncertainty about when and how it will adapt. Although the ultimate goal for human-AI teams should be dynamic adaptation, a training mode that permits team members to control the agent's adaptation manually while they learn when and to what levels it adapts will help teams overcome this in-trepidation faster and more successfully.

Teams are fluid constructs and gain and lose membership, and so another important aspect of this training mode is that it could be utilized to introduce a single member of the team to the agent, even after the team as a whole has become comfortable with its AI teammate(s). Most professions, such as computer security (Killcrece, Kossakowski, Ruefle, & Zajicek, 2003), require members to obtain certifications relevant to their job role, and one way to ensure that all members of a team have trained with the AI teammate in training mode could be to attach a certification to it or make it part of the employee on-boarding process. Just as new employees must complete a series of tasks and training modules when they are hired, training with an AI teammate in its training mode should be one of the on-boarding requirements.

### 5.4. Limitations and future work

One major limitation to this study is that everything the participants considered was hypothetical and abstract. Numerous studies have shown that people have a tendency to be more risk adverse when their own assets are at stake and less so in hypothetical situations (Norwood & Jayson, 2007). Future research that utilizes an experiment to implement some of these scenarios and then ask the participants their comfort levels would increase the validity of these findings. Another limitation of the research is the strictly one-agent scenarios. People tend to become less comfortable as the number of AI systems increase [? ?], and scenarios in which there are more than one adaptive AI teammate may change how autonomous human teammates want them to be, as well as if that makes the delineation of when and how autonomous they should be less clear. Further research that incorporates larger, more diverse human-AI teammates will provide more realistic insights on the research questions. Another area for future research exposed in this study is the concern over explainability as an AI agent's autonomy level changes. Many of out participants expressed that as an agent becomes more independent is becomes more important for it to be able to explain its decisions and actions to the rest of the team, as at the end of the day it will be the human beings, not the AI, held responsible for the consequences of those actions. Future research should address the changing explainability requirements for adaptive autonomous agents, and how that affects teammate perceptions of competent and trustworthy AI teammates.

## 6. Conclusion

The unique capabilities of modern day AI have increased the desire for the use of AI agents in the professional work space (Reim, Åström, & Eriksson, 2020). As AI agents take on full team roles and become teammates, it would be advantageous for them to possess higher levels of autonomy (O'Neill et al., 2020). The issue is that in complex work environments, such as cyber incident response, there may be various levels at which you want such an agent to operate. This study explored the design and cognitive effects of adaptive autonomous agents in a human-AI teaming context, an area yet to be studied in human-AI literature. We studied the design of adaptive AI teammates for CSIRTs utilizing a factorial survey and qualitative interviews. Insights from these studies show that professional teams that have a well defined team work cycle, such as the incident response cycle, can link the right

LOA for an AI teammate to specific points in those processes, thus enabling dynamic adaptation. Higher LOAs should exist in situations where the agent's actions are more predictable and the team has more experience. Agents should operate at lower LOAs in situations where the correct procedures are ill-defined and there is less time for the agent to convey its decisions. In approaching the design of these adaptable AI agents, the more dynamic they are in their adaptation the more likely they will be accepted and treated as legitimate teammates, as opposed to tools used by the team.

## CRediT authorship contribution statement

**Allyson I. Hauptman:** Conceptualization, Methodology, Investigation, Writing – original draft. **Beau G. Schelble:** Methodology, Data curation, Formal analysis, Writing – original draft. **Nathan J. McNeese:** Supervision, Writing – review & editing. **Kapil Chalil Madathil:** Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


Ahmad, Atif, Desouza, Kevin C, Maynard, Sean B, Naseer, Humza, & Baskerville, Richard L (2020). How integration of cyber security management and incident response enables organizational learning. *Journal of the Association for Information Science and Technology*, *71*(8), 939–953.

Alufaisan, Yasmeen, Marusich, Laura R, Bakdash, Jonathan Z, Zhou, Yan, & Kantarcioglu, Murat (2020). Does explainable artificial intelligence improve human decision-making? arXiv preprint arXiv:2006.11194.

Armstrong, Robert L. (1987). The midpoint on a five-point likert-type scale. *Perceptual and Motor Skills*, *64*(2), 359–362.

Benítez, Isabel, Van de Vijver, Fons, & Padilla, José Luis (2019). A mixed methods approach to the analysis of bias in cross-cultural studies. *Sociological Methods & Research*, Article 0049124119852390.

Blair, Erik (2015). A reflexive exploration of two qualitative data coding techniques. *Journal of Methods and Measurement in the Social Sciences*, *6*(1), 14–29.

Braun, Virginia, & Clarke, Victoria (2012). Thematic analysis.

Burke, Anthony (2020). Robust artificial intelligence for active cyber defence. *Alan Turing Institute, Tech. Rep*.

Burkle, Frederick M. (1996). Acute-phase mental health consequences of disasters: implications for triage and emergency medical services. *Annals of Emergency Medicine*, *28*(2), 119–128.

Charmaz, Kathy (2006). Constructing grounded theory: A practical guide through qualitative analysis. sage.

Cooke, Nancy J, Gorman, Jamie C, Myers, Christopher W, & Duran, Jasmine L (2013). Interactive team cognition. *Cognitive Science*, *37*(2), 255–285.

Cooke, Nancy J., Gorman, Jamie C., & Winner, Jennifer L. (2007). Team cognition.

De Lucia, Michael J., Newcomb, Allison, & Kott, Alexander (2019). Features and operation of an autonomous agent for cyber defense. arXiv preprint arXiv:1905.05253.

Demir, Mustafa, & Cooke, Nancy J. (2014). Human teaming changes driven by expectations of a synthetic teammate. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 16–20.

Donevski, Michael, & Zia, Tanveer (2018). A survey of anomaly and automation from a cybersecurity perspective. In *2018 IEEE Globecom workshops (GC Wkshps)* (pp. 1–6). IEEE.

Fereidunian, Alireza, Lehtonen, Matti, Lesani, Hamid, Lucas, Caro, & Nordman, Mikael (2007). Adaptive autonomy: smart cooperative cybernetic systems for more humane automation solutions. In *2007 IEEE International conference on systems, man and cybernetics* (pp. 202–207). IEEE.

Fiske, Amelia, Henningsen, Peter, & Buyx, Alena (2020). The implications of embodied artificial intelligence in mental healthcare for digital wellbeing. In *Ethics of digital well-being* (pp. 207–219). Springer.

Gavin, Helen (2008). Thematic analysis. *Understanding Research Methods and Statistics in Psychology*, 273–282.

de Greef, Tjerk, & Arciszewski, Henryk (2008). Combining adaptive automation and adaptive teams in a naval command centre. In *Proceedings of the 15th European conference on cognitive ergonomics: the ergonomics of cool interaction* (pp. 1–4).

Harvey-Jordan, Stephanie, & Long, Sarah (2001). The process and the pitfalls of semi-structured interviews. *Community Practitioner*, *74*(6), 219.

Hasan, Raza, Raghav, Akshyadeep, Mahmood, Salman, & Hasan, M Asim (2011). Artificial intelligence based model for incident response. In *2011 International conference on information management, innovation management and industrial engineering, vol. 3* (pp. 91–93). IEEE.

Hernandez, Alejandro S, Karimova, Tahmina, Nelson, Douglas H, Ng, E, Nepal, B, & Schott, E (2017). Mission engineering and analysis: innovations in the military decision making process. In *Proceedings of the american society for engineering management (ASEM) 2017 International annual conference: reimagining systems engineering and management* (pp. 521–530). ASEM.

Hox, Joop J., Kreft, Ita G. G., & Hermkens, Piet L. J. (1991). The analysis of factorial surveys. *Sociological Methods & Research*, *19*(4), 493–510.

Irons, Alastair, & Lallie, Harjinder Singh (2014). Digital forensics to intelligent forensics. *Future Internet*, *6*(3), 584–596.

Jain, Ritu, & Suman, Ugrasen (2015). A systematic literature review on global software development life cycle. *ACM SIGSOFT Software Engineering Notes*, *40*(2), 1–14.

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

Jarrett, Aaron, & Choo, Kim-Kwang Raymond (0000). The impact of automation and artificial intelligence on digital forensics, *Wiley Interdisciplinary Reviews: Forensic Science*, e1418,

Kerstholt, José, Barnhoorn, Jonathan, Hueting, Tom, & Schuilenborg, Lotte (2018). Automation as an intelligent teammate: Social psychological implications. In *NATO-HAT Symposium on human autonomy teaming*.

Killcrece, Georgia, Kossakowski, Klaus-Peter, Ruefle, Robin, & Zajicek, Mark (2003). *State of the practice of computer security incident response teams (CSIRTs): Technical Report*, CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST.

Kim, Se Yun (2021). Group affect and group cohesion in human-agent teams.

Klien, Glen, Woods, David D, Bradshaw, Jeffrey M, Hoffman, Robert R, & Feltovich, Paul J (2004). Ten challenges for making automation a" team player" in joint human-agent activity. *IEEE Intelligent Systems*, *19*(6), 91–95.

Liu, Han, Lai, Vivian, & Tan, Chenhao (2021). Understanding the effect of out-of-distribution examples and interactive explanations on human-AI decision making. arXiv preprint arXiv:2101.05303.

Liu, Zheng, Li, Xiaohan, You, Zeyu, Yang, Tao, Fan, Wei, & Yu, Philip (2021). Medical triage chatbot diagnosis improvement via multi-relational hyperbolic graph neural network. In *Proceedings of the 44th International ACM SIGIR conference on research and development in information retrieval* (pp. 1965–1969).

Love, Jonathon, Selker, Ravi, Marsman, Maarten, Jamil, Tahira, Dropmann, Damian, Verhagen, Josine, et al. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1–17.

Lyn Paul, Celeste, Blaha, Leslie M, Fallon, Corey K, Gonzalez, Cleotilde, & Gutzwiller, Robert S (2019). Opportunities and challenges for human-machine teaming in cybersecurity operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 442–446.

Mahaini, Mohamad Imad, Li, Shujun, & Sağlam, Rahime Belen (2019). Building taxonomies based on human-machine teaming: Cyber security as an example. In *Proceedings of the 14th International conference on availability, reliability and security* (pp. 1–9).

McGee, Ethan T., & McGregor, John D. (2016). Using dynamic adaptive systems in safety-critical domains. In *Proceedings of the 11th International symposium on software engineering for adaptive and self-managing systems* (pp. 115–121).

McNeese, Nathan J, Demir, Mustafa, Cooke, Nancy J, & Myers, Christopher (2018). Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors*, *60*(2), 262–273.

McNeese, Nathan J, Schelble, Beau G, Canonico, Lorenzo Barberis, & Demir, Mustafa (2021). Who/what is my teammate? Team composition considerations in human-AI teaming. arXiv preprint arXiv:2105.11000.

Mepham, Kevin, Louvieris, Panos, Ghinea, Gheorghita, & Clewley, Natalie (2014). Dynamic cyber-incident response. In *2014 6th International conference on cyber conflict (CyCon 2014)* (pp. 121–136). IEEE.

Mitropoulos, Sarandis, Patsos, Dimitrios, & Douligeris, Christos (2006). On incident handling and response: A state-of-the-art approach. *Computers & Security*, *25*(5), 351–370.

Moro, Christina, Lin, Shayne, Nejat, Goldie, & Mihailidis, Alex (2019). Social robots and seniors: a comparative study on the influence of dynamic social features on human–robot interaction. *International Journal of Social Robotics*, *11*(1), 5–24.

Musick, Geoff, Zhang, Rui, McNeese, Nathan J, Freeman, Guo, & Hridi, Anurata Prabha (2021). Leveling up teamwork in esports: Understanding team cognition in a dynamic virtual environment. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–30.

Nikolaidis, Stefanos, & Shah, Julie (2013). Human-robot cross-training: computational formulation, modeling and evaluation of a human team training strategy. In *2013 8th ACM/IEEE International conference on human-robot interaction* (pp. 33–40). IEEE.

Norwood, F. Bailey, & Jayson, L. (2007). Forecasting hypothetical bias: A tale of two calibrations. In *Environmental economics, experimental methods* (pp. 469–487). Routledge.

Nyre-Yu, Megan M. (2019). *Determining system requirements for human-machine integration in cyber security incident response* (Ph.D. thesis), Purdue University Graduate School.

Nyre-Yu, Megan, Gutzwiller, Robert S., & Caldwell, Barrett S. (2019). Observing cyber security incident response: qualitative themes from field research. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 437–441.

Oh, Changhoon, Song, Jungwoo, Choi, Jinhan, Kim, Seonghyeon, Lee, Sungwoo, & Suh, Bongwon (2018). I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. In *Proceedings of the 2018 CHI Conference on human factors in computing systems* (pp. 1–13).

O'Neill, Thomas, McNeese, Nathan, Barron, Amy, & Schelble, Beau (2020). Human–autonomy teaming: A review and analysis of the empirical literature. *Human Factors*, Article 0018720820960865.

Parasuraman, Raja, Sheridan, Thomas B., & Wickens, Christopher D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *30*(3), 286–297.

Pinto-Llorente, Ana María, Sánchez-Gómez, M Cruz, & Pedro Costa, António (2020). Qualitative and mixed methods researches in social sciences. In *Eighth international conference on technological ecosystems for enhancing multiculturality* (pp. 193–196).

Reim, Wiebke, Åström, Josef, & Eriksson, Oliver (2020). Implementation of artificial intelligence (AI): a roadmap for business model innovation. *AI*, *1*(2), 180–191.

Rist, Thomas, André, Elisabeth, Baldes, Stephan, Gebhard, Patrick, Klesen, Martin, Kipp, Michael, et al. (2004). A review of the development of embodied presentation agents and their application fields. *Life-Like Characters*, 377–404.

Rouse, William B. (1988). Adaptive aiding for human/computer control. *Human Factors*, *30*(4), 431–443.

Rouse, William B., & Rouse, Sandra H. (1983). *A framework for research on adaptive decision aids*: *Technical Report*, ALPHATECH INC BURLINGTON MA.

Scott, Cliff, & Medaugh, Melissa (2017). Axial coding. *The International Encyclopedia of Communication Research Methods*, *10*, Article 9781118901731.

Seeber, Isabella, Bittner, Eva, Briggs, Robert O, De Vreede, Triparna, De Vreede, Gert-Jan, Elkins, Aaron, et al. (2020). Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, *57*(2), Article 103174.

Smith, Carol J. (2019). Designing trustworthy AI: A human-machine teaming framework to guide development. arXiv preprint arXiv:1910.03515.

Staves, Alex, Balderstone, Harry, Green, Benjamin, Gouglidis, Antonios, & Hutchison, David (2020). A framework to support ICS cyber incident response and recovery. In *The 17th International conference on information systems for crisis response and management*.

Suzanne Barber, K., Goel, Anuj, & Martin, Cheryl E. (2000). Dynamic adaptive autonomy in multi-agent systems. *Journal of Experimental & Theoretical Artificial Intelligence*, *12*(2), 129–147.

Tambe, Milind, Pynadath, David V, Chauvat, Nicholas, Das, Abhimanyu, & Kaminka, Gal A (2000). Adaptive agent integration architectures for heterogeneous team members. In *Proceedings fourth international conference on multiagent systems* (pp. 301–308). IEEE.

Taylor, Brian J. (2006). Factorial surveys: Using vignettes to study professional judgement. *British Journal of Social Work*, *36*(7), 1187–1207.

Van Schaik, Paul, & Ling, Jonathan (2003). Using on-line surveys to measure three key constructs of the quality of human–computer interaction in web sites: psychometric properties and implications. *International Journal of Human-Computer Studies*, *59*(5), 545–567.

Wang, Dakuo, Weisz, Justin D, Muller, Michael, Ram, Parikshit, Geyer, Werner, Dugan, Casey, et al. (2019). Human-ai collaboration in data science: Exploring data scientists' perceptions of automated ai. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24.

Wen, Senhao, Rao, Yu, & Yan, Hanbing (2018). Information protecting against APT based on the study of cyber kill chain with weighted Bayesian classification with correction factor. In *Proceedings of the 7th International Conference on Informatics, Environment, Energy and Applications* (pp. 231–235).

Wickens, Christopher D, Li, Huiyang, Santamaria, Amy, Sebok, Angelia, & Sarter, Nadine B (2010). Stages and levels of automation: An integrated meta-analysis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *54*(4), 389–393.

Yeo, M Lisa, Rolland, Erik, Ulmer, Jackie Rees, & Patterson, Raymond A (2014). Risk mitigation decisions for IT security. *ACM Transactions on Management Information Systems (TMIS)*, *5*(1), 1–21.

Zhang, Rui, McNeese, Nathan J., Freeman, Guo, & Musick, Geoff (2021). " An ideal human" expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3), 1–25.