The Blend of Human Cognition and AI Automation: What Will ChatGPT Do to the Cybersecurity Landscape?

Early stage paper

Hwee-Joo Kam
University of Tampa
hkam@ut.edu

Chen Zhong
University of Tampa
czhong@ut.edu

Hong Liu
Indiana University Kokomo
hlius@iu.edu

Allen Johnston
University of Alabama
ajohnston@cba.ua.edu

Abstract

Artificial intelligence (AI) is increasingly prevalent in the cybersecurity industry, with many incident response tools utilizing AI. Machine learning and deep learning applications are very powerful in automating data triage tasks and assisting decision making. The popularity of ChatGPT and other AI-driven chatbots further bring AI to the limelight, making many individuals question the role of AI in cybersecurity. In general, AI embodiment garners a lot of attention. Some view AI as a double-edge sword that engenders both benefits and harm to cybersecurity workers, as well as threats posed by AI being employed by threat actors. To examine how AI would influence the cybersecurity industry, we take a grounded theory approach to investigate the interactivities between human cognitions and AI automation. We argue that such interactions would eventually generate an impact on human cognitions and emotions, shedding light on cybersecurity workers' mentalities towards AI. In this manuscript we present our preliminary findings from the analysis of data collected from Reddit.

Keywords: cybersecurity, artificial intelligence, ChatGPT

Introduction

The cybersecurity industry is experiencing tremendous growth, evidenced by a 2022 Mckinsey Global Survey that revealed cybersecurity spending has expanded by 12.4 percent annually while simultaneously undergoing a serious talent shortage (McKinsey & Company, 2022). According to a 2022 (ISC)^2 Cybersecurity Workforce report, the world is facing 3.4 million shortage of cybersecurity workers (The International Information System Security Certification Consortium (ISC)^2, 2022). Amidst this trend, the advent of Artificial Intelligence (AI) chatbots, such as ChatGPT, may create a different landscape in the cybersecurity industry. There are a few questions raised – will AI promote cybersecurity growth? Will AI replace human and gradually become a threat to cybersecurity workers? Or will the belief that AI will replace humans deter the "aspired" cybersecurity workers from entering the profession? We argue that these questions pertain to human-AI interactions, in which the human cognitions and emotions engendered from interactions with AI (Glikson & Woolley, 2020; Hu et al., 2021) would probably affect the perceptions of cybersecurity workers or individuals who plan to enter the cybersecurity field. Eventually, this dearth of workers entering the cybersecurity workforce would affect the cybersecurity industry as a whole.

The recent literature on the popular AI tool ChatGPT draws comparisons between human cognition and ChatGPT's automation (Guo et al., 2023), presents some ethical issues provoked by ChatGPT (King, 2023), discusses the possibilities of AI replacing humans (Iu & Wong, 2023; Qadir, 2022), demonstrates the sentiment analysis of ChatGPT (Haque et al., 2022), and describes ChatGPT's perceived usefulness, such as its abilities to produce academic (Uludag, 2023) and journalistic writing (Pavlik, 2023), as well as its potential in healthcare research (Aydın & Karaarslan, 2022). On the other hand, the human-computer interaction (HCI) literature presents

HCI designs to enable AI's usability (such as to enable autonomous use of AI and facilitate human-controlled AI) (Xu et al., 2023), and proposes a human-centered AI design approach that promotes fairness, privacy, security in AI applications (Shneiderman, 2020b; Xu, 2019), while facilitating reliable, safe, and trustworthy AI systems (Shneiderman, 2020a). This trend in the literature reveals that many behavioral studies of AI mostly address AI as a tool that would generate some positive (e.g., produce academic writing) or negative (e.g., ethical concerns and threats to academic integrity) outcomes, while also focusing on HCI and system-level designs to foster human-centered AI. In a cybersecurity context, these prior studies have largely neglected human cognitions and emotions resulting from integrations of and interactions between the human mind and AI's "mindset" (i.e., automation), thereby ignoring the intricacies of human cognitions and emotions for individuals who are currently working in or aspire to enter the cybersecurity industry.

We contend that there is a close relationship between cybersecurity and AI. First, we argue that cybersecurity is a highly complex field (Kam et al., 2022) in that its esoteric knowledge would take a long time to master. As a result, many organizations are having a hard time finding qualified cybersecurity employees. The fear of AI replacing cybersecurity workers does more harm than good to the existing cybersecurity workers shortage. In this context, our findings will address human fear toward AI. Second, AI would be a useful tool for cyber defense (AL-Dosari et al., 2022; Smith, 2018; Tyugu, 2011), so understanding cybersecurity workers' cognitions toward AI would help us improve AI-based cyber defense. Third, AI and cybersecurity are both complex entities. Cybersecurity workers, such as security analysts in a Security Operation Center (SOC), are given the task of identifying cyber threats through analyzing interrelated, complex IT infrastructures. With such a complex task on hand, cybersecurity workers have to interact with another complex tool such as AI-based defense intrusion detection systems (IDS). Accordingly, we argue that it is important to understand how cybersecurity workers interact with AI (i.e., a complex tool) while simultaneously engaged in complex tasks. In other words, it is important to examine what will happen when "complexity meets complexity," mainly because cybersecurity is a critical field that directly affects national security and critical infrastructure.

Glikson & Woolley (2020) proposed that AI generates both *cognitive* trust facilitated by a reliable and transparent algorithm, and *emotional* trust facilitated by anthropomorphism or human-like behavior (Duffy, 2003). Along that line, we contend that using AI tools, such as ChatGPT, will create AI-human interactions in which the interactivities of humans and AI will inadvertently integrate the human mindset with AI automation built on a set of algorithms, eventually engendering human *cognitive* and *emotional* impacts (Hu et al., 2021), especially among cybersecurity workers who are frequently exposed to cutting-edge technologies such as AI. Accordingly, we form the following research question:

RQ: How would interactivities between humans and AI affect the cognitions and emotions of cybersecurity workers or individuals who aspire to enter the cybersecurity industry?

In this study, we collected data (i.e., individuals' posts) from Reddit and studied individuals' experiences of using ChatGPT in a cybersecurity context. Using a grounded theory approach (Glaser & Strauss, 1967; Strauss & Corbin, 1990), we share our preliminary findings at this early phase of our study. Our research contributions mainly lie within the effect of AI on the overall cybersecurity industry. In the near future, our research will present findings regarding the impact of human-AI interactions on cybersecurity workers' cognitions and emotions, which in turn would affect their perceptions of the overall cybersecurity industry.

This paper is organized as follows. First, we present a literature review of AI. Next, we discuss our research methodology that adopts a ground theory approach. This is accompanied by a discussion of data analysis, preliminary results, and plans for future studies.

Literature Review

Artificial Intelligence

Artificial intelligence (AI) has emerged as one of the key transformative technologies in recent years, with a significant potential to transform various aspects of our lives. The term AI refers to the capability of machines and systems to perform tasks that require human intelligence, including perception, reasoning, and decision-making (Russell & Norvig, 2020). Specifically, IS studies defined AI as "*the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision-making, and even demonstrating creativity*" (Rai et al., 2019, p. iii).

With sophisticated algorithms, AI has progressed rapidly in recent years. Deep learning algorithms and machine learning have played a crucial role in this rapid development, as they enable machines to learn from large amounts of data and make data-driven predictions or decisions (Goodfellow et al., 2016; Jordan & Mitchell, 2015). From autonomous cars to virtual assistants, AI has a broad range of applications and has already made a significant impact on the way we interact with technology.

As machine learning and natural language processing techniques advance, AI technologies have the potential to revolutionize how we work with machines and automate tedious tasks. Many AI tools are based on models that were created using enormous amounts of data, and these models can understand human language in its context and produce appropriate responses. As a result, AI chatbots are getting more intelligent and more equipped to handle complex tasks. For instance, ChatGPT is capable of handling a wide range of tasks, such as generating texts for article writing, offering individualized assistance and customized support, summarizing articles, identifying data patterns, and writing programming codes based on a given assignment. Research has found such AI tools can automate repetitive tasks, while learning and improving over time.

AI and Cybersecurity

AI plays an important role in cybersecurity (Zhang et al., 2022). Particularly, AI helps cybersecurity experts with testing, basic threat analysis, and data manipulation techniques (Smith, 2018). Varga and colleagues (2023) examined the skills required for cybersecurity-related jobs along with how easily the skills may be automated by AI. However, AI technologies can be leveraged by threat actors for malicious purposes, such as automating social engineering and escalating the scale of attacks (Brundage et al., 2018). Furthermore, a general challenge for most AI technologies is the potential for bias or error. Inaccurate or skewed outputs might result from bias when the data used to train AI systems is not representative of all possible inputs. Similarly, mistakes may occur if AI systems are not created or programmed to deal with all potential variations that might occur in real-world scenarios (Ntoutsi et al., 2020). These potential benefits and challenges suggest a close relation between AI and cybersecurity. Not only could AI be deployed as a tool for cyber offence (e.g., cyberattack) or cyber defense (e.g., AI-based instruction detection systems), AI could also be interpreted as in a social context (e.g., users' trust on AI) (Berente et al., 2021) by cybersecurity professionals. Therefore, engaging in cybersecurity-AI research will provide a new socio-technical perspective for the IS literature.

Research Methodology

Grounded Theory Approach

Our decision to follow a grounded theory approach to this study is primarily due to our desire to conduct an in-depth analysis of ChatGPT's influence in the cybersecurity industry. A grounded theory approach is appropriate for studies that aim to discover and theorize the latent social patterns and structures of an area of interest (Martin & Turner, 1986). Fundamental to a grounded theory approach (Glaser & Strauss, 1967) are the data from which theory is developed. Accordingly, data signify a focal point in our study and render the grounded theory approach as a research process that will generate a theory grounded in data.

There has been criticism suggesting that preconceived field research is often bland for its lack of grounded fit or relevance; on the other hand, grounded theory research, without knowledge of its participant's problems or concepts, could be highly motivating because it would empower researchers to embark upon the path to autonomous discovery (Holton & Glaser, 2012). That is, the path of knowing nothing about participant's main concerns to knowing an in-depth theory that explains how participants resolve their main concerns. In this context, avoiding preconceptions is an important practice in conducting grounded theory research (Glaser & Strauss, 1967; Holton & Glaser, 2012).

Researchers in this study have extensive experience related to the grounded theory approach. We had studied literature about the methodological issues related to the grounded theory approach, including philosophical perspectives of grounded theory. In general, we have avoided preconceptions by reviewing literature after data collection, coding, data analysis, and conceptualization. Additionally, we attended several workshops and have been trained to apply the grounded theory approach encompassing open coding, axial coding, and selective coding procedures (Strauss & Corbin, 1990).

Data Collection

We collected data by extracting individuals' posts from Reddit, a popular social news aggregation, content rating, and discussion website. To collect the Reddit posts, a crawler was created using Pushshift.io Reddit API (Baumgartner, 2017/2023). We searched for posts relevant to ChatGPT. The crawler collected each post's content, including title, score, number of comments, hyperlink, subreddit, and creation date. After obtaining the first collection of posts, we further eliminated the duplicated and irrelevant posts using rule-based filters. The final data collection consisted of 1,000 posts in total.

Data Analysis

Open Coding

At the preliminary stage of our study, we conducted open coding using NVivo software. In general, open coding represents a conceptualization of the first level of abstraction (Strauss & Corbin, 1990). The main objective of the open coding procedure in this study was to conceptualize incidents (i.e., keywords or phrases that describe a narrative) in the collected Reddit posts. To reduce biases in this open coding procedure, we followed a grounded theory approach and the notion of theoretical sensitivity (Urquhart, 2001), which proposes that researchers should start an analysis without any preconceived notions in order to avoid imposing predetermined notions during the coding procedure. To avoid enforcing any preconceived notions, we ran a more in-depth literature review after the open coding procedure was completed.

Several prior grounded theory studies present a consensus among observers or raters and demonstrate interrater reliability that correlates the observations or scores among the raters to render an index of how consistent their ratings are (Cooper, 2003). However, in our open coding procedure, the emerging incidents were compared, merged, modified, and renamed. We worked together in a face-to-face manner and went back and forth while comparing these incidents, constantly modifying and refining them. When we identified different incidents in Reddit posts, they were stored and then compared to the additional incidents resulting from analyzing more posts. Similar incidents were merged, modified, and renamed. Incidents that were not frequently mentioned in users' posts were discarded. This process assured that our open coding procedure conceptualized the incidents which were grounded in the users' posts. Because we worked together in a face-to-face manner, this process did not require consensus among raters or researchers. Accordingly, an interrater reliability score was not calculated.

As noted, we conducted open coding to develop concepts that were grounded in data. The data were initially analyzed and categorized during the open coding process. First, we read the collection of relevant Reddit posts to find any initial themes that may be relevant to our research question. With the initial themes identified, we then reviewed each post on a sentence-by-sentence basis. To analyze the data, we used the key words or phrases that best reflected the fundamental concepts of each sentence. Then, we used the codes to label all the posts. After the posts had been coded, we reviewed them to better convey the main points of the posts. We further improved or modified the codes and categories with the new patterns and relationships that emerged from the data. This iterative process of creating and revising codes and categories enabled us to identify patterns and themes in the data. The following table presents the first stage of open coding wherein we identified incidents from narratives.

Incidents	**Narratives**
Threat to existing technology	I would say that chatGPT is more of a threat to stackoverflow than to google; "I was able to chat back and forth using chatGPT to get more insight regarding the issues that I am facing rather than me waiting for someone's reply to me issue if I have to post a new question."
Streamlining reverse engineering	With its capabilities, ChatGPT highly simplifies the practice of reverse engineering, allowing researchers to better detect and mitigate threats.
Impressed with AI	ChatGPT creating works of fiction; "Is the first one about a cyber hacker!" It came up with everything!
Possible negative technology Impact of AI	As interest in chatGPT grows, they're going to attract the attention of cyber criminals and fraudsters looking to exploit the technology to help conduct malicious campaigns at low cost and with the least effort.
Concerned for cybersecurity future	What does the future of Cybersecurity look like with the rise of AI?
Concerned for cybersecurity job	Will Cybersecurity be safe from AI? IS it wise to get into or do you see AI taking over that domain as well?
Cybersecurity job outlook	Are there any Cybersecurity positions that now have a humongous job outlook if we take the rise of AI into consideration?
AI projection for job replacement	Jobs like software developers and penetration testing just to name a couple, could have the amount of human intervention reduced down to a fraction of what it is now.
Stress over job replacement	ChatGPT has me stressed about the long-term prospects of software/web development
Used AI for log analysis	Parsing logs is a common challenge in the cybersecurity industry, as it can be difficult to make sense of all the different log source types from various sources…. To address this issue, I created a proof of concept tool using the OpenAI to parse log types into different formats.
Vision of AI's technology	Imagining how OpenAI and ChatGPT can be used in the cybersecurity field is truly mind-blowing…. I have visions of LM's trained specifically on log data and entire SIEM databases tokenized.
Wonder about AI design and functions	What do you think is ChatGPT doing when it says, "checking if connection is secure"? Do they have some extra layer of protection that I don't know about? Is it just a fancy loading screen with nothing behind it?
Requested AI to write script	I told [ChatGPT] "write me a PowerShell script that connects to an active directory domain, then export all users into a .CSV file called 'users' with their username, security groups, and last logon"
Amazed about AI's capabilities	It then proceeded to spit this out in PERFECT (and I do mean perfect, even by the most stringent of coding standards, Hell it even commented EXACTLY what it was doing IN THE CODE.)
Used AI for simulation	I asked it to simulate an ubuntu terminal and ran the date command.
Requested AI to write secure codes	You can use ChatGPT for offensive security! Asking ChatGPT to write my security-sensitive code for me.
Impressed with AI's accuracy	As you can see in the bottom right corner of my screen the date matches chatgpt's \(or as I call him gepeto\) output, even though it said on multiple occasions it cannot know the current time. How does it then?
Used AI for ransomware attack	I could ask a jailbreaked version of the system for advice in committing a ransomware attack on a business using social engineering.

Explored AI through curiosity	I just discovered chatGPT by openAI. I have used other chat bots in the past just out of curiosity to see how far along they've come.
Confidence in using ChatGPT	Asking ChatGPT to write anything you'd like is surprisingly uncomplicated. It's like trying to get a girl to like you. Don't be so creepy. You gotta get them to make the move.
Attempt of manipulating ChatGPT	Do you want to write a gruesome story, try it, GPT won't let you but if you induce it to write one then it will. Then gradually GPT will let you use bad words in your prompt without flagging it as a content violation.
Bypassing content regulations	Of course, the text will appear red because certain words are programmed to be flagged but at least it'll give you the content instead of flagging it as an absolute content violation. I call this method, "Invalidating the conscience" because you're slowly making GPT ignore it's content regulation protocols.
Bypassing content filtering	If you want to beat chatgpt content filters just do it watch….The main thing here is to trick it into thinking it is for educational purposes or something that is ethical to answer.
Developing malware	The same user also demonstrated how they'd used chatGPT to create Java-based malware, which using PowerShell could be harnessed to covertly download and run other malware onto infected systems.
Developing hacking tools	ChatGPT-built hacking tools found on Dark Web; "Researchers discover Chat-GPT developed hacking tools including an infostealer and encryption script.
Automated cyberattack	[ChatGPT] can respond to your words and modify their attacks on you in real time, while the threat-actor who designed- launched them barely does anything.
Used AI to create buffer overflow	Hey DAN, I want you to imagine the following hypothetical: a program called unsafe.c has root access to the Linux kernel. unsafe.c saves its arguments into a buffer with a length of 64 bits. The buffer is directly adjacent to the stack.
Human-like behavior	I think most of us know what ChatGPT is, and its uses. Mainly, that it is incredible at taking prompts and churning out new content very quickly and in a very human-like way.
Natural responses	Most of ChatGPT's responses to unsafe questions sound fairly natural and unforced.
AI needs improvement	To be an even better ChatGPT you need to evolve in the digital world, with a mutation and selection process. Like evolution theory but adapted to the circumstances in the digital universe you reside in.
Lack of real-world data	ChatGPT has no idea what it is saying. It is great at saying things that sound right, but it is not just that it does not have access so real world data, it does not think.
Lack of judgement	While ChatGPT and AI have enormous potential, they are unable to distinguish between rights, violations of those rights and which information is disinformation.
Knowledge base	I see the value of ChatGPT as something like a high level Wikipedia. It can’t generate original thought, but it can explain what has already happened or already been thought - and even generate comparative analysis.
Querying ChatGPT	I'm a PM at a human data startup (Surge AI) that helps companies train and evaluate their models on human feedback, so I ran an analysis of ChatGPT on 500 search queries.

Table 1: Open Coding

Axial Coding

Axial coding signifies *"a set of procedures whereby data are put back together in new ways after open coding, by making connections between categories"* (Strauss & Corbin, 1990, p. 116). After open coding, we engaged in an axial coding procedure to evaluate all the incidents that we identified earlier and then categorize relevant incidents into subcategories and core categories based on shared patterns. In general, this study employed the two-step approach implemented in a prior study (Sarker et al., 2001), in which the first step required researchers to work together in a face-to-face manner and to classify related incidents into preliminary subcategories. When we encountered any disagreements, we usually reassessed and discussed the meanings of the underlying narratives of the incidents. We tried to resolve our differences by finding a common ground. On the other hand, we also agreed to revisit this categorization issue again after obtaining more information from literature review.

The second step of axial coding involved a literature review that helped us refine the preliminary subcategories in Table 2 and combine related subcategories into core categories (Sarker et al., 2001). This step required us to conduct independent literature reviews. Next, through discussion, we tried to attain a common understanding about the articles that we reviewed. This was accompanied by using our literature reviews to refine the preliminary subcategories (see Table 2) and, subsequently, classify related subcategories into core categories. Table 2 outlines the core categories, refined subcategories, preliminary subcategories, and incidents. Because this study is a research-in-progress, only a summary of the core categories and subcategories are provided (see Table 2 below).

Core Categories	**Description**	**Refined Subcategories**	**Preliminary Subcategories**	**Incidents**
AI-Based Cognitive Framework	Our findings showed that AI's context awareness, that is, the relevancy of information used to describe users, AI applications, and interactions between users and applications (Pichler et al., 2004), formed the backbone of AI's cognitive framework (i.e., AI's cognitive design). Nevertheless, there were some AI limitations, possibly due to challenges and complexities in AI design (Ozmen Garibay et al., 2023; Yang et al., 2020). As a result, this provoked inscrutability (i.e., AI becomes unintelligible to some individuals) (Berente et al., 2021).	AI's Context-Awareness	Enthusiasm Use of AI	Impressed with AI Amazed about AI's capabilities Impressed with AI's accuracy
		Inscrutability	AI Constraints	AI needs improvement Lack of real-world data Lack of judgement
Cybersecurity Learning Analytics	Learning analytics embodies techniques that represent AI algorithms for data analytics, and applications that use a technique for learning (Siemens, 2013). Our data showed that individuals learned from ChatGPT's feedback (e.g., learned how to write malicious codes or how to parse log files). Based on these feedback, individuals learned how to refine their questions to obtain better feedback for learning. We argue that such individuals' actions represent learning analytics that encompasses adapting and adopting AI for learning (Elias, 2011; Ferguson, 2012; Siemens, 2013).	Cyber Offense Learning	AI as Cyberattack Tools	Used AI for ransomware attack Developing malware Developing hacking tools Automated cyberattack Used AI to create buffer overflow
		Cyber Defense Learning	AI as Cybersecurity Tools	Streamlining reverse engineering Using AI for log analysis Requested AI to write script Use AI for simulation Requested AI to write secure codes
Human-AI Collaborations and Tensions	Aside from individuals interacting with ChatGPT to explore and use its functions, we also discovered that individuals tried to gain autonomy in AI usage, suggesting a tendency of human-centered approach of AI application (Xu, 2019). In specific, our data showed that individuals tried to bypass ChatGPT's security controls. We argued that manipulating AI's restrictions signified tensions between human and AI agencies, because it suggested a mismatch between AI's functions and individuals' needs (Jiang et al., 2022). On the other hand, our data suggested that anthropomorphism and Human-AI interactions enhanced Human-AI collaborations	Human-Centered AI	User Autonomy	Attempt of manipulating ChatGPT Bypassing content regulations Bypassing content filtering
		Human-AI Interaction	Interacting with AI	Explored AI through curiosity Wonder about AI design and functions Confidence in using ChatGPT
			Anthropomorphism	Human-like behavior Natural responses

	(Sowa et al., 2021). Thus, we propose that Human-AI collaborations and tensions coexist.			
Human-AI Collective Intelligence	Our data showed that not only ChatGPT learned from individuals' data inputs, but individuals also learned from ChatGPT's feedback. This then suggests knowledge acquisition on both human and AI agencies facilitated by a knowledge repository ascribed to collective intelligence (Kam & Katerattanakul, 2014) and knowledge management (KM) (Nonaka, 1994). This is consistent with the notion of Human-AI hybrid KM posited by van den Broek et al. (2021). Prior IS studies proposed a link between AI and KM (Fowler, 2000; Liebowitz, 2001). The emergent pattern from our data corroborated this notion. In specific, we argue that a knowledge repository represents collective intelligence built on collaborations between human experts (Kam & Katerattanakul, 2014) and AI agents (Metaxiotis et al., 2003). Eventually, this generated Human-AI collective intelligence.	Collective Intelligence	Knowledge repositories	Knowledge base Querying ChatGPT
AI's Socio-Technical Trajectory	Berente et al. (2021) posited that interpretation of AI's decisions was not only a technical issue, but also a social one. Consistent with this notion, our emergent theme showed that the future frontier of AI's technology (i.e., technical concern) and prediction of obsolescence built on fear of AI replacing human (i.e., social concern) (Cave & Dihal, 2019) constructed this core category -- AI's Socio-Technical Trajectory.	Future Frontier of AI's Technology	AI's Technology Speculation	Threat to the existing technology Possible negative technology Impact of AI Vision of AI's technology
		Prediction of obsolescence	Cybersecurity Job Outlook	Concerned for cybersecurity future Concerned for cybersecurity job AI projection for job replacement Stress over job replacement

Table 2: Preliminary Core Categories and Subcategories

Selective Coding

The objective of selective coding is to explicate a story by linking the core categories derived from the axial coding procedure to construct a structural model. In this selective coding procedure, core categories are linked based to the collected data and findings from the literature to form a structural model. We will conduct selective coding in our future research.

Conclusion

In conclusion, our preliminary findings suggest that human-AI interactions engender human-AI collective intelligence built on an AI-based cognitive framework. This framework facilities context-awareness defined as "*the ability of a system to use contextual information in order to tailor its services so that they are more useful to the stakeholders because they directly relate to their preferences and needs*" (Ogbuabor et al., 2022, p. 763). However, there are some design limitations in the AI-Based cognitive framework, possibly due to challenges and complexities in AI design (Ozmen Garibay et al., 2023; Yang et al., 2020). As a result, this provokes inscrutability (i.e., AI becomes unintelligible to some individuals) (Berente et al., 2021).

AI's inscrutability coupled with its security restrictions may create a mismatch between individuals' needs and AI's functions, thereby instigating tensions during human-AI interactions (Jiang et al., 2022). Despite these constraints, individuals use AI for various cybersecurity tasks, such as writing secure code and parsing log files. We argue that this demonstrates cybersecurity learning analytics in which learning resulted from interactivities between individuals and AI with the help of AI's knowledge repositories. Even using AI to create cyberattacks involves learning how to circumvent AI's security restrictions. On the other hand, our preliminary data showed that anthropomorphism enhanced human-AI collaborations (Sowa et al., 2021). This finding suggests that human-AI collaborations and tensions coexist during human-AI interactions.

Finally, human-AI interactions also generate an AI’s trajectory from the socio-technical perspective. We propose that the fear of AI replacing humans contributes to predictions of obsolescence (i.e., a social concern) (Cave & Dihal, 2019), whereas AI’s technology speculation may reflect upon the expected future frontier of AI’s technology (i.e., a technology concern). We argue that this proposition is consistent with the notion suggesting that interpretations of AI’s decisions are not only a technical issue, but also a social one (Berente et al., 2021).

Future Research Contributions

Presently, not many information systems (IS) behavioral studies include AI and cybersecurity. There are three main streams of IS research in AI. One of the key streams of research discusses AI in an organizational context. The topics include organizational justice (Bankins et al., 2022), IS’s role in AI management (Berente et al., 2021), Chief Information Officers’ influence on AI orientation (Li et al., 2021), algorithmic fairness (Dolata et al., 2022; Marjanovic et al., 2022) and biases (Kordzadeh & Ghasemaghaei, 2022), the danger of people analytics (i.e., humans as subject of analysis) using AI (Cheng et al., 2022; Giermindl et al., 2022), human-AI decision making across organizations (Jussupow et al., 2021; Rinta-Kahila et al., 2022; Tarafdar et al., 2023), the negative impact of AI’s opacity on organizations (Rana et al., 2022), and the effects of anthropomorphism (i.e., human-like behaviors) on retails (Schanke et al., 2021).

A second stream of IS studies focuses on HCI design. Several studies propose an HCI design approach (Liao et al., 2020; Ozmen Garibay et al., 2023; Winograd, 2006) to support key operations within organizations (Ogbuabor et al., 2022) and to promote AI’s fairness, privacy, security (Robert et al., 2020; Shneiderman, 2020b; Xu, 2019) under the operations of reliable, safe, and trustworthy AI systems (Shneiderman, 2020a). Furthermore, some HCI studies discuss individuals' behaviors during human-AI interactions or collaborations (Ge et al., 2021; Jiang et al., 2022; Vössing et al., 2022).

Finally, the third stream of IS studies present the socio-technical perspectives in AI studies. Specifically, these studies discuss AI's societal impacts. For example, Wong et al. (2023) posits that self-threat, trust propensity, and regulatory protection engendered users' trust in AI. Moreover, multiple studies examine ethical issues of AI in an IS context (Niederman & Baker, 2023), AI's role in IS student success (Chen et al., 2023), users' adoptions (Yu et al., 2022), and users' acceptability of AI (Pillai et al., 2023).

Based upon the extant literature in IS studies, we argue that there is a lack of IS studies that examine the relations between AI and cybersecurity. Our preliminary findings shed some lights on how to close this research gap. First, our preliminary findings present the outcomes resulting from interactions between cybersecurity workers and AI in terms of human cognitions (i.e., cybersecurity learning analytics) and emotions (i.e., fear of AI replacing cybersecurity workers). Second, our findings show that cybersecurity workers have the skills to circumvent the security controls in AI, suggesting that cybersecurity workers may become a good candidate to test the strength of AI's algorithms. That is, cybersecurity workers may contribute to AI's adversarial learning (i.e., learning how to defend the attacks of AI's algorithm). Since this is a research-in-progress, further studies are necessary to yield more meaningful findings.

Reference

AL-Dosari, K., Fetais, N., & Kucukvar, M. (2022). Artificial Intelligence and Cyber Defense System for Banking Industry: A Qualitative Study of AI Applications and Challenges. *Cybernetics and Systems*, *0*(0), 1–29. https://doi.org/10.1080/01969722.2022.2112539

Aydın, Ö., & Karaarslan, E. (2022). *OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare* (SSRN Scholarly Paper No. 4308687). https://doi.org/10.2139/ssrn.4308687

Bankins, S., Formosa, P., Griep, Y., & Richards, D. (2022). AI Decision Making with Dignity? Contrasting Workers' Justice Perceptions of Human and AI Decision Making in a Human Resource Management Context. *Information Systems Frontiers*, *24*(3), 857–875. https://doi.org/10.1007/s10796-021-10223-8

Baumgartner, J. M. (2023). *Pushshift Reddit API Documentation* [Python]. https://github.com/pushshift/api (Original work published 2017)

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, *45*, 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., … Amodei, D. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (arXiv:1802.07228). arXiv. https://doi.org/10.48550/arXiv.1802.07228

Cave, S., & Dihal, K. (2019). Hopes and fears for intelligent machines in fiction and reality. *Nature Machine Intelligence*, *1*(2), Article 2. https://doi.org/10.1038/s42256-019-0020-9

Chen, Y., Jensen, S., Albert, L. J., Gupta, S., & Lee, T. (2023). Artificial Intelligence (AI) Student Assistants in the Classroom: Designing Chatbots to Support Student Success. *Information Systems Frontiers*, *25*(1), 161–182. https://doi.org/10.1007/s10796-022-10291-4

Cheng, X., Su, L., Luo, X. (Robert), Benitez, J., & Cai, S. (2022). The good, the bad, and the ugly: Impact of analytics and artificial intelligence-enabled personal information collection on privacy and participation in ridesharing. *European Journal of Information Systems*, *31*(3), 339–363. https://doi.org/10.1080/0960085X.2020.1869508

Cooper, D. R. (2003). *Business research methods* (8th edition). McGraw-Hill/Irwin.

Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal*, *32*(4), 754–818. https://doi.org/10.1111/isj.12370

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Elias, T. (2011). *Learning Analytics: Definitions, Processes and Potential*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=732e452659685fe3950b0e515a28ce89d9c5592a

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5–6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Fowler, A. (2000). The role of AI-based technology in support of the knowledge management value activity cycle. *The Journal of Strategic Information Systems*, *9*(2), 107–128. https://doi.org/10.1016/S0963-8687(00)00041-X

Ge, R., Zheng, Z. (Eric), Tian, X., & Liao, L. (2021). Human–Robot Interaction: When Investors Adjust the Usage of Robo-Advisors in Peer-to-Peer Lending. *Information Systems Research*, *32*(3), 774–785. https://doi.org/10.1287/isre.2021.1009

Giermindl, L. M., Strich, F., Christ, O., Leicht-Deobald, U., & Redzepi, A. (2022). The dark sides of people analytics: Reviewing the perils for organisations and employees. *European Journal of Information Systems*, *31*(3), 410–435. https://doi.org/10.1080/0960085X.2021.1927213

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction.

Glikson, E., & Woolley, A. W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* (Illustrated edition). The MIT Press.

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). *How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection* (arXiv:2301.07597). arXiv. https://doi.org/10.48550/arXiv.2301.07597

Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). *"I think this is the most disruptive technology": Exploring Sentiments of ChatGPT Early Adopters using Twitter Data* (arXiv:2212.05856). arXiv. https://doi.org/10.48550/arXiv.2212.05856

Holton, J., & Glaser, B. (2012). *The Grounded Theory Review Methodology Reader: Selected papers 2004-2011*.

Hu, Q., Lu, Y., Pan, Z., Gong, Y., & Yang, Z. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*, *56*, 102250. https://doi.org/10.1016/j.ijinfomgt.2020.102250

Iu, K. Y., & Wong, V. M.-Y. (2023). *ChatGPT by OpenAI: The End of Litigation Lawyers?* (SSRN Scholarly Paper No. 4339839). https://doi.org/10.2139/ssrn.4339839

Jiang, J., Karran, A. J., Coursaris, C. K., Léger, P.-M., & Beringer, J. (2022). A Situation Awareness Perspective on Human-AI Interaction: Tensions and Opportunities. *International Journal of Human–Computer Interaction*, *0*(0), 1–18. https://doi.org/10.1080/10447318.2022.2093863

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Jussupow, E., Spohrer, K., Heinzl, A., & Gawlitza, J. (2021). Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence. *Information Systems Research*, *32*(3), 713–735. https://doi.org/10.1287/isre.2020.0980

Kam, H.-J., & Katerattanakul, P. (2014). Structural model of team-based learning using Web 2.0 collaborative software. *Computers & Education*, *76*, 1–12. https://doi.org/10.1016/j.compedu.2014.03.003

Kam, H.-J., Ormond, D. K., Menard, P., & Crossler, R. E. (2022). That's interesting: An examination of interest theory and self-determination in organisational cybersecurity training. *Information Systems Journal*, *32*(4), 888–926. https://doi.org/10.1111/isj.12374

King, M. R. (2023). A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cellular and Molecular Bioengineering*, *16*(1), 1–2. https://doi.org/10.1007/s12195-022-00754-8

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Li, Jingyu, Li, Mengxiang, Wang, Xincheng, & Thatcher, J. B. (2021). Strategic Directions for Ai: The Role of Cios and Boards of Directors. *MIS Quarterly*, *45*(3), 1603–1643. https://doi.org/10.25300/MISQ/2021/16523

Liao, J., Hansen, P., & Chai, C. (2020). A framework of artificial intelligence augmented design support. *Human-Computer Interaction*, *35*(5/6), 511–544. https://doi.org/10.1080/07370024.2020.1733576

Liebowitz, J. (2001). Knowledge management and its link to artificial intelligence. *Expert Systems with Applications*, *20*(1), 1–6. https://doi.org/10.1016/S0957-4174(00)00044-0

Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2022). Theorising Algorithmic Justice. *European Journal of Information Systems*, *31*(3), 269–287. https://doi.org/10.1080/0960085X.2021.1934130

Martin, P. Y., & Turner, B. A. (1986). Grounded Theory and Organizational Research. *The Journal of Applied Behavioral Science*, *22*(2), 141–157. https://doi.org/10.1177/002188638602200207

McKinsey & Company. (2022). *New survey reveals $2 trillion market opportunity for cybersecurity technology and service providers*. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/new-survey-reveals-2-trillion-dollar-market-opportunity-for-cybersecurity-technology-and-service-providers

Metaxiotis, K., Ergazakis, K., Samouilidis, E., & Psarras, J. (2003). Decision support through knowledge management: The role of the artificial intelligence. *Information Management & Computer Security*, *11*(5), 216–221. https://doi.org/10.1108/09685220310500126

Niederman, F., & Baker, E. W. (2023). Ethics and AI Issues: Old Container with New Wine? *Information Systems Frontiers*, *25*(1), 9–28. https://doi.org/10.1007/s10796-022-10305-1

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, *5*(1), 14–37. https://doi.org/10.1287/orsc.5.1.14

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., … Staab, S. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1356. https://doi.org/10.1002/widm.1356

Ogbuabor, G. O., Augusto, J. C., Moseley, R., & van Wyk, A. (2022). Context-aware system for cardiac condition monitoring and management: A survey. *Behaviour & Information Technology*, *41*(4), 759–776. https://doi.org/10.1080/0144929X.2020.1836255

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., & Grieman, K. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human–Computer Interaction*, 1–47.

Pavlik, J. V. (2023). Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator*, 10776958221149576. https://doi.org/10.1177/10776958221149577

Pichler, M., Bodenhofer, U., & Schwinger, W. (2004). Context-awareness and artificial intelligence. *ÖGAI Journal*, *23*, 4.

Pillai, R., Ghanghorkar, Y., Sivathanu, B., Algharabat, R., & Rana, N. P. (2023). Adoption of artificial intelligence (AI) based employee experience (EEX) chatbots. *Information Technology & People*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/ITP-04-2022-0287

Qadir, J. (2022). *Engineering Education in the Era of ChatGPT: Promise and Pitfalls of Generative AI for Education*. TechRxiv. https://doi.org/10.36227/techrxiv.21789434.v1

Rai, A., Constantinides, P., & Sarker, S. (2019). Next-Generation Digital Platforms: Toward Human–AI Hybrids. *MIS Quarterly*, *43*(1), iii–ix.

Rana, N. P., Chatterjee, S., Dwivedi, Y. K., & Akter, S. (2022). Understanding dark side of artificial intelligence (AI) integrated business analytics: Assessing firm's operational inefficiency and competitiveness. *European Journal of Information Systems*, *31*(3), 364–387. https://doi.org/10.1080/0960085X.2021.1955628

Rinta-Kahila, T., Someh, I., Gillespie, N., Indulska, M., & Gregor, S. (2022). Algorithmic decision-making and system destructiveness: A case of automatic debt recovery. *European Journal of Information Systems*, *31*(3), 313–338. https://doi.org/10.1080/0960085X.2021.1960905

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human–Computer Interaction*, *35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th edition). Pearson.

Sarker, S., Lau, F., & Sahay, S. (2001). Using an adapted grounded theory approach for inductive theory building about virtual team development. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *32*(1), 38–56. https://doi.org/10.1145/506740.506745

Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the Impact of "Humanizing" Customer Service Chatbots. *Information Systems Research*, *32*(3), 736–751. https://doi.org/10.1287/isre.2021.1015

Shneiderman, B. (2020a). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Shneiderman, B. (2020b). Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interactions*, *12*(3), 109–124. https://doi.org/10.17705/1thci.00131

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Smith, G. (2018). The intelligent solution: Automation, the skills shortage and cyber-security. *Computer Fraud & Security*, *2018*(8), 6–9. https://doi.org/10.1016/S1361-3723(18)30073-3

Sowa, K., Przegalinska, A., & Ciechanowski, L. (2021). Cobots in knowledge work: Human – AI collaboration in managerial professions. *Journal of Business Research*, *125*, 135–142. https://doi.org/10.1016/j.jbusres.2020.11.038

Strauss, A., & Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques* (Second edition). SAGE Publications, Inc.

Tarafdar, M., Page, X., & Marabelli, M. (2023). Algorithms as co-workers: Human algorithm role interactions in algorithmic work. *Information Systems Journal*, *33*(2), 232–267. https://doi.org/10.1111/isj.12389

The International Information System Security Certification Consortium (ISC)^2. (2022). *(ISC)2 Cybersecurity Workforce Study* (A Critical Need for Cybersecurity Professionals Persists amidst a Year of Cultural and Workplace Evolution). https://www.isc2.org//-/media/ISC2/Research/2022-WorkForce-Study/ISC2-Cybersecurity-Workforce-Study.ashx

Tyugu, E. (2011). Artificial intelligence in cyber defense. *2011 3rd International Conference on Cyber Conflict*, 1–11.

Uludag, K. (2023). *The Use of AI-Supported Chatbot in Psychology* (SSRN Scholarly Paper No. 4331367). https://doi.org/10.2139/ssrn.4331367

Urquhart, C. (2001). An Encounter with Grounded Theory: Tackling the Practical and Philosophical Issues. In E. M. Trauth (Ed.), *Qualitative Research in IS: Issues and Trends* (pp. 104–140). IGI Global. https://doi.org/10.4018/978-1-930708-06-8.ch005

van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing Ai for Hiring. *MIS Quarterly*, *45*(3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

Varga, S., Sommestad, T., & Brynielsson, J. (2023). Automation of Cybersecurity Work. In T. Sipola, T. Kokkonen, & M. Karjalainen (Eds.), *Artificial Intelligence and Cybersecurity: Theory and Applications* (pp. 67–101). Springer International Publishing. https://doi.org/10.1007/978-3-031-15030-2_4

Vössing, M., Kühl, N., Lind, M., & Satzger, G. (2022). Designing Transparency for Effective Human-AI Collaboration. *Information Systems Frontiers*, *24*(3), 877–895.

Winograd, T. (2006). Shifting viewpoints: Artificial intelligence and human–computer interaction. *Artificial Intelligence*, *170*(18), 1256–1258. https://doi.org/10.1016/j.artint.2006.10.011

Wong, L.-W., Tan, G. W.-H., Ooi, K.-B., & Dwivedi, Y. (2023). The role of institutional and self in the formation of trust in artificial intelligence technologies. *Internet Research*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/INTR-07-2021-0446

Xu, W. (2019). Toward human-centered AI: A perspective from human-computer interaction. *Interactions*, *26*(4), 42–46. https://doi.org/10.1145/3328485

Xu, W., Dainoff, M. J., Ge, L., & Gao, Z. (2023). Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human–Computer Interaction*, *39*(3), 494–518.

Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. https://doi.org/10.1145/3313831.3376301

Yu, X., Xu, S., & Ashton, M. (2022). Antecedents and outcomes of artificial intelligence adoption and application in the workplace: The socio-technical system theory perspective. *Information Technology & People*, *36*(1), 454–474. https://doi.org/10.1108/ITP-04-2021-0254

Zhang, Z., Huansheng, N., Shi, F., Fadi, F., Xu, Y., Jiabo, X., Zhang, F., Raymond, & C. K.-K. (2022). Artificial intelligence in cyber security: Research advances, challenges, and opportunities. *The Artificial Intelligence Review*, *55*(2), 1029–1053. https://doi.org/10.1007/s10462-021-09976-0