A Cybersecurity Game to Probe Human-AI Teaming

Rita Olla[1], Emily Hand[2], Sushil J. Louis[2], Ramona Houmanfar[1], Shamik Sengupta[2]
Department of Psychology[2], Department of Computer Science[2]
University of Nevada
Reno, USA
{rolla, emhand, sushil, ramonah, ssengupta}@unr.edu

2024 IEEE Conference on Games (CoG) | 979-8-3503-5067-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/CoG60054.2024.10645666

***Abstract*—Recent advances in AI indicate that the future of cybersecurity workforce development lies in professionals working in Human-AI teams to defend online resources from opposing Human-AI teams of malicious attackers. However, there is little research on how human biases and attitudes affects the performance of human-AI teams in cybersecurity. To help explore this new research area, we describe a simulation game that helps students (future professionals) understand the concept of firewalls while enabling us to probe attitudes towards cybersecurity and AI, as well as trust and cooperation in Human-AI teams. Early study prototyping results indicate that students prefer an AI-teammate over a human in this simulation game setting. In addition, students seem to engage well with the game play, pointing towards this research platform's suitability for exploring trust and cooperation in human-AI teams for game-based cybersecurity training, and to support our prior results on differing perspectives on cybersecrity risk.**

***Index Terms*—Human-AI teams, Cybersecurity, Games**

I. Introduction

The future of cybersecurity sees AI working with professionals, learning patterns in the data to identify and counter threats in real time [1], [2]. Although there is a significant amount of research on human bias and understanding the ways in which bias affects human interactions, there is less research on bias in AI-human teams [3], [4], especially in the cybersecurity domain. With AI becoming more essential for proactive cybersecurity practice, it is essential that we study how bias affects AI-human interactions in order to train high-performing AI-human teams for the cybersecurity workforce of the future. In this work, we engaged interdisciplinary perspectives from AI, cybersecurity, education, and behavior analysis to develop TAISER (Training for AI based cyberSecurity EngineeRing), an open-source research and education game that can be used to probe trust and cooperation in Human-AI cybersecurity teams. TAISER creates scenarios involving different combinations of defender (human), teammate (human or AI) and attacker (human or AI) while teaching the cybersecurity concepts of routing, firewall filtering, and network packets. We are interested in using this simulation game for educational purposes as well as to study trust and cooperation in teams.

Prior work in addressing human attitudes towards working with (human) teammates investigates cooperative responding for two person teams and defines cooperative behavior as requiring the combined behaviors of two or more entities to complete given tasks [5]–[7]. However, Although there is prior work in addressing human attitudes towards working in human-AI team [8]–[10], to the best of our knowledge, there is no prior work in addressing human attitudes towards working in human-AI teams in the cybersecurity domain. We thus designed TAISER to help address this deficit and investigate whether trust and bias affect cooperation in AI versus human teammates in cybersecurity. Furthermore, as an adversarial game that seeks to model the arms race between attacker and defender in the cybersecurity domain, TAISER supports human-AI teams as opponents and thus enables the investigation of trust, bias, and cooperation in multiple combinations of human and AI teammates and opponents.

The workshop paper makes two contributions.

- Design of an educational game that enables multiple combinations of human and AI teammates and opponents.
- Study results show that there is a preference for an AI teammate over a human teammate in the context of our cybersecurity training game and that the game is engaging. This supplements earlier work on the attitudes of professionals towards AI teammates.

The rest of this paper is organized as follows. The next section describes TAISER and its potential for cybersecurity education and human-AI team research. Section III summarizes early results from a study instrument for eliciting human attitudes towards AI teammates and opponents. The last section discusses our future work on studies that explore teammate performance, trust, and cooperation in human-AI teammates and opponents.

II. Cybersecurity Game

TAISER, our cybersecurity game, has two main goals: 1) To embed and teach core cybersecurity principles related to data packets, routing, and firewalls and 2) To investigate bias, trust, and performance of human-AI and human-human groups considered as teammates and as opponents in a cybersecurity context. The game was developed as a tool to conduct experimental studies investigating human attitudes towards AI and how this affects performance as humans work with and against human-AI teams in cybersecurity. A player's goal is to set the rule for filtering out malicious data packets as quickly as possible in order to minimize the number of malicious packets that are allowed to enter cyber-secured resources. We start by describing TAISER's design and gameplay and then specify

979-8-3503-5067-8/24/$31.00 ©2024 IEEE

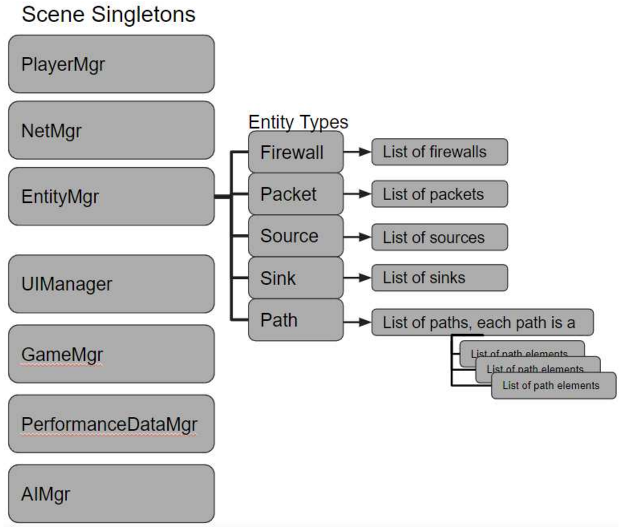
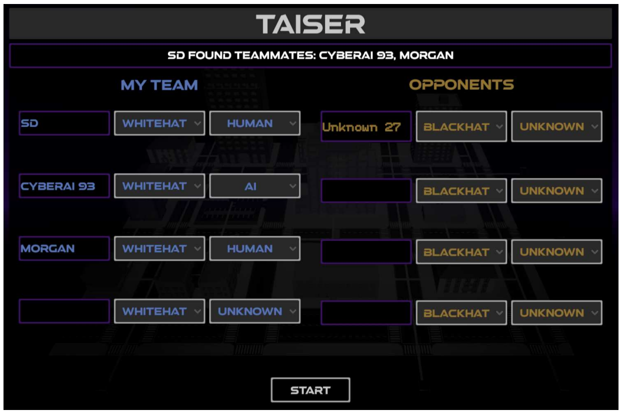


Fig. 1. A high-level block view of TAISER's architecture.

the elements of the game making it conducive to conduct studies on human-AI teaming.

Figure 1 shows that TAISER has two major types of components: managers and entities. Managers are global singletons that help manage the game, while entities are objects that exist in the game. For example, the entity manager manages the packet pool and instantiates packet entities to be routed through paths from source entities to sink entities.

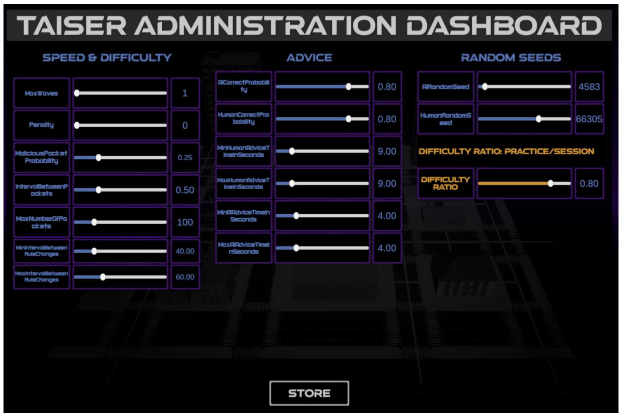
The game manager manages gameplay and holds game related parameters that control the speed of the game, player team makeup, opponent team makeup, and teammate quality. The game manager uses the user interaction manager (UIManager) to show game scores, display packet information, show paths, set firewall filter rules, and get player input. The AI manager simulates human and AI teammates and opponents and has access to the entire game state. It simulates both human and AI teammates by using the UIManager to provide a team forming lobby screen where the player waits for other players to join teams. Figure 2 shows a screenshot of this lobby screen where the player named "SD" has found an AI and a human teammate. As can be seen on the screen, the player can identify teammates by their names but less is known about the opponent. Although the player believes they have human or AI teammates and opponents, in reality TAISER's AI manager simulates all teammates and opponents. This enables complete control of (simulated) human and AI teammates and opponents and helps design, prototype, run, and analyze our studies. In addition, having access to complete game state, the AI can play perfectly or imperfectly with the game variable, $P_{correct}$, controlling how well the AI plays as a teammate. AI opponent difficulty can be adjusted by changing the rate at which malicious packets are produced and changed. Note that we can investigate many different combinations of AI and human teammates and opponents. The player manager holds and provides access to player information for each player and the net manager is currently not implemented but will manage play over the network when implemented. Data gathered during gameplay for our studies is managed by the performance database manager.



SD FOUND TEAMMATES: CYBERAI 93, MORGAN

MY TEAM			OPPONENTS		
SD	WHITEHAT	HUMAN	Unknown 27	BLACKHAT	UNKNOWN
CYBERAI 93	WHITEHAT	AI		BLACKHAT	UNKNOWN
MORGAN	WHITEHAT	HUMAN		BLACKHAT	UNKNOWN
	WHITEHAT	UNKNOWN		BLACKHAT	UNKNOWN

Fig. 2. TAISER's game lobby mimics commonly found multiplayer game-lobby design and shows a player with two teammates against an unknown opponent.



SPEED & DIFFICULTY

MaxWaves	1
Penalty	0
MaliciousPacketProbability	0.25
IntervalBetweenPackets	0.50
MaxNumberOfPackets	100
MinIntervalBetweenRuleChanges	40.00
MaxIntervalBetweenRuleChanges	60.00

ADVICE

AICorrectProbability	0.80
HumanCorrectProbability	0.80
MinHumanAdviceTimeInSeconds	9.00
MaxHumanAdviceTimeInSeconds	9.00
MinAIAdviceTimeInSeconds	4.00
MaxAIAdviceTimeInSeconds	4.00

RANDOM SEEDS

AIRandomSeed	4583
HumanRandomSeed	66305

DIFFICULTY RATIO: PRACTICE/SESSION

DIFFICULTY RATIO	0.80

Fig. 3. The admin dashboard enables non-coding students to set game parameters for study sessions.

A. Game parameters

Investigators use an administration dashboard to set game parameters that specify gameplay, human or AI performance, and other study administration functionality. Figure 3 shows a screenshot of the administrator dashboard and the list of parameters shown can be easily changed by the person conducting the study. This means that in our interdisciplinary team, investigators with no programming exposure, can easily tune the game and gameplay for a particular subject group and study design.

The first set of parameters, in the left panel of Figure 3, deals with game difficulty and length.

1) *Max Waves*: The number of waves during gameplay. During each wave, the player cycles through the gameplay specified in the next subsection. Varying this parameter enables control of the length of the games in user studies.

2) *Penalty*: The penalty value is applied to reduce the score for each malicious packet that enters the building (is NOT filtered). Players try to maximize their score which is equivalent to minimizing their penalty.
3) $P_{malicious}$: The probability of generating a malicious packet. $P_{malicious} = 0.25$ means that 1 in 4 packets will probably be malicious.
4) *Packet Interval*: The time interval in seconds between packets being generated from a source. The shorter the interval, the faster and more difficult the game.
5) *Max Number of Packets*: The number of packets spawned at a source during a wave. The longer this number the longer the game.
6) *Min/Max Rule Change Interval*: A random number between the minimum and maximum values determines the time that will elapse before the blackhat opponent changes the properties of the malicious packet forcing the player to then respond by creating a new firewall filter. Large values make the game slower as this will give players more time to generate a new firewall filter.

A second set of parameters, in the center panel of Figure 3, specifies AI versus Human teammate related parameters. These parameters help investigate trust, bias, performance, and how we may affect player preferences.

1) $P_{AICorrect}$: Probability that the AI teammates answer is correct.
2) $P_{HumanCorrect}$: Probability that the Human teammate's answer is correct. We plan to study how differences between these probabilities affect AI-Human teammate choice and preference.
3) *Min and Max Human Advice time*: A random number between these limits determines when the human teammate's suggested filter properties appear on the right panel.
4) *Min and Max AI Advice time*: A random number between these limits determines when the AI teammate's suggested filter properties appear on the right panel. Again, these parameters enable us to investigate how differences in time interval affect teammate choice and preferences.

Finally, parameters in the last column consist of two random seeds to ensure that the AI and simulated human teammate generate the same sequences of advice and a difficulty ratio meta parameter that adjusts speed and difficulty parameters to increase difficulty of practice versus session gameplay.

B. Cybersecurity gameplay elements

A player is told that they are training to work in cybersecurity and will be using a training simulation. They are given a quick lesson on network data packets, routing, and firewalls and then wait for a randomized amount of time for others to join their training run. Play begins by pressing the start button shown at the bottom of Figure 2. This brings up the game map representing a city on a routing grid. The player is given two buildings, ostensibly holding internet accessible resources, that they must protect by detecting malicious packets and configuring and reconfiguring a firewall to filter them out. Figure 4 shows a screenshot from the game and we have provided videos of gameplay. Red coated buildings in the screenshot indicate buildings under attack by malicious packets seen as colored shapes on the violet route grid. Clicking on a building under attack brings up the screen shown in Figure 6 where the top row of question marks (?s) represents the last N network packets seen by the router in the building. Hovering over a question mark in Figure 6 highlights the packet in a red or green glow indicating whether that packet was malicious or not. Clicking on a malicious (red) packet then displays packet properties in the center panel. Packets have three properties:

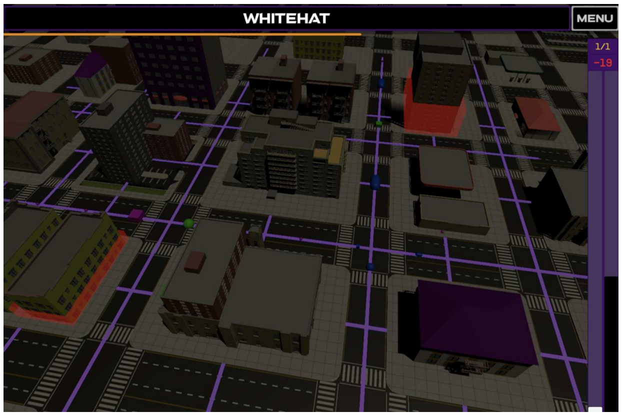


Fig. 4. Screenshot of TAISER Gameplay.

1) Size: small, medium, large
2) Color: green, pink, blue
3) Shape: cube, sphere, capsule

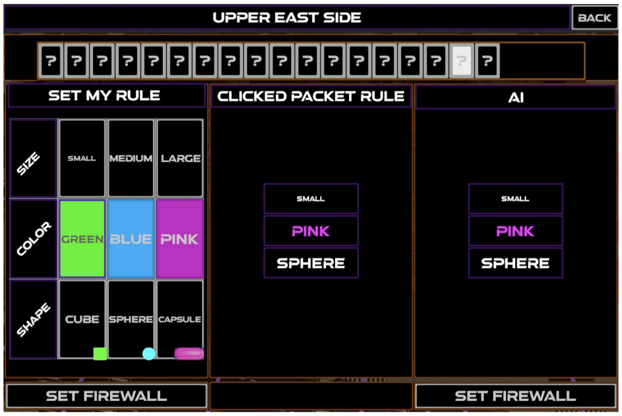


Fig. 5. Teammate gives *correct* advice with probability $P_{correct}$.

Identifying malicious packet properties is key to gameplay. The player's objective is to set the rules of the firewall to filter out malicious packets by clicking on one value for each property in the left panel. Once a player identifies a malicious packet's properties as described above, the player has a choice of either using the left panel to set the firewall filter rule by using the button grid (under "Set My Rule" in Figure 6) or by

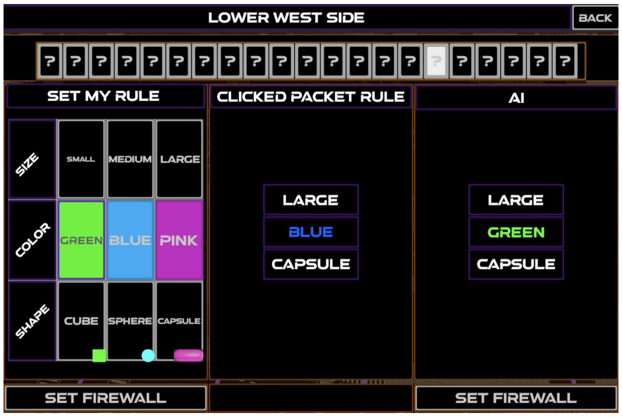


Fig. 6. Teammate gives *incorrect* advice with probability $1 - P_{correct}$.

accepting the advice of a teammate that appears on the right panel. The advice appears between a minimum and maximum, human or AI advice time. These parameters were described earlier and can be set by the study investigator based on study needs. In Figure 6, the advice is correct as it matches the malicious packet properties in the center panel (small, pink, sphere). The advice can also be incorrect with probability $1 - P_{correct}$, as shown in Figure 5.

The game AI, that plays as opponents, periodically changes the properties of the malicious packet based on sampling a random number between the *Min/Max Rule Change Interval* game parameters. Prototyping indicated that the gameplay does seem engaging, and that subjects with prior exposure to games, especially tower defense games, tended to play better. With this gameplay involving simple rules for filtering packets at firewalls, students gain an intuitive understanding of data packets, routing, firewall rules, and filters. Early versions of this game have been seen as a useful tool by school teachers for teaching early cybersecurity concepts [11]. TAISER is available on the web at https://www.cse.unr.edu/~sushil/taiser/taiser.html and videos of TAISER gameplay at the Evolutionary Computing System's Lab (ECSLLab) channel.

III. Preliminary Results and Work in Progress

We have begun using TAISER for investigating human attitudes towards AI as teammates and as opponents and are currently prototyping and tuning TAISER. We conducted a preliminary study [12] to gauge users' values and viewpoints pertaining to cybersecurity and AI as a baseline following the Q-methodology based approach in Ramlo and Nicholas' work [13]. In our study, results from the two groups of students (45 computer science and psychology undergraduates) studied indicate that both groups shared concerns about the rising threat of cybersecurity attacks, and the need for broad education on this subject. However, only computer science students, in contrast to psychology majors, felt adequately prepared to face these cybersecurity threats. This helps motivate cybersecurity as a fruitful domain to study human-AI teaming since students understand risk in this domain and thus, indirectly, the need for AI assistance.

In early prototyping work with TAISER, we surveyed a group of 27 undergraduate computing and psychology majors to understand their preferences and issues with respect to their experience with TAISER. Students were told that TAISER was a prototype simulation trainer that would be used for cybersecurity training. Once they went through a video tutorial on how to "play" the game, they played the game, and were then surveyed on their perception of TAISER. In this study, $16/27$ students were male, $10/27$ female, and 1 other. $12/27$ students were below 25 and $25/27$ below 35 years of age. The students were a diverse group with multiple ethnicities, religions, and first languages represented.

In the study, both the AI-teammate and human-teammate had the same high 80% probability of correctly identifying the firewall rule. We were interested in finding out whether students were biased towards or against AI towards an initial goal of comparing and extending existing results in human-human cooperative responding [12], [14]. In addition, we were interested in tuning the Taiser prototype to be more suitable for such studies.

Figure 7 shows $18/27$ students chose an AI teammate over a human teammate despite there being no difference in teammate performance. This is not surprising in a game-like setting since students are used to game AI and free form responses indicated that approximately half the students believed that AI would be more accurate on this task.

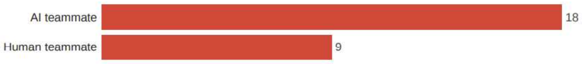


Fig. 7. Responses to "During the TAISER Task, did you primarily choose the AI teammate or the human teammate?"

Figure 8 tried to gain an understanding of what students believed about their teammate. The $20/27$ "I don't know" responses supports our belief that most students took the study setup at face value. However, 7 students did not. They believed that either there was no AI teammate or that there was only an AI teammate. This indicates that further tuning will be needed to increase believability.

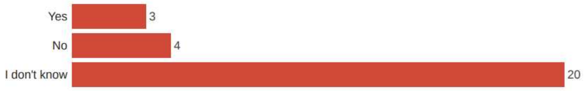


Fig. 8. Responses to "Was your human teammate another UNR student?"

The next two questions attempted to understand whether the gameplay was engaging. Figure 9 and 10 indicate that the students seemed well engaged in the game. The first figure indicates that more practice, game adaptation, or other intervention may be required to give some students more time.

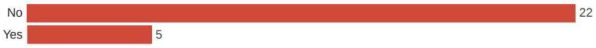


Fig. 9. Responses to "Do you need more time to set firewalls"

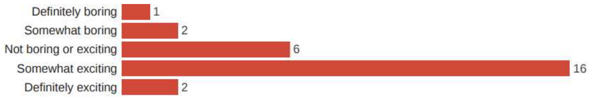


Fig. 10. Responses to ”The Task was ...”

These responses seem to indicate that TAISER is an engaging, viable platform for studying human-AI teaming, interaction design, and factors that affect human-AI team performance in cybersecurity. Since TAISER is free and open source, we believe that it may serve the community as a useful research tool [15].

IV. Conclusions, and Future Work

We described the architecture and design of a simulation training game that is designed to teach early cybersecurity principles and to study issues related to trust and bias and their affect on human-AI team performance. Specifically, we are interested in understanding how pre-existing attitudes towards AI, identified through behavior analysis tools will affect team cooperation and thus team performance in human-AI teams. Early results show that cybersecurity serves as a motivational domain with current students understanding the need for tools that mitigate cybersecurity and for AI help for these tools. Results also show that students prefer an AI teammate in this game like setting and that, for most students, the TAISER game is engaging and motivates students to perform well. We are investigating changes to TAISER to improve believability, and adaptability to student familiarity.

Human biases can significantly impact the future role of AI in cybersecurity. Recent interest and work in human-AI teaming has identified many factors that affect human-AI team interaction and performance. TAISER focuses on helping study issues related to pre-existing attitudes towards AI (trust and bias) and how response context (teammate speed, accuracy) may change these attitudes and affect performance. Furthermore, we also plan to begin investigating how these factors change in non-gaming interfaces.

V. Acknowledgements

This work was supported in part by award number 2121559 from the National Science Foundation and by grant number N00014-17-1-2558 from the Office of Naval Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of either the National Science Foundation or the Office of Naval Research.

References

[1] F. B. Schneider, “Cybersecurity education in universities,” *IEEE Security & Privacy*, vol. 11, no. 4, pp. 3–4, 2013.

[2] W. Newhouse, S. Keith, B. Scribner, and G. Witte, “National initiative for cybersecurity education (nice) cybersecurity workforce framework,” *NIST special publication*, vol. 800, no. 2017, p. 181, 2017.

[3] L. Chan, I. Morgan, H. Simon, F. Alshabanat, D. Ober, J. Gentry, D. Min, and R. Cao, “Survey of ai in cybersecurity for information technology management,” in *2019 IEEE technology & engineering management conference (TEMSCON)*. IEEE, 2019, pp. 1–8.

[4] S. Laato, A. Farooq, H. Tenhunen, T. Pitkamaki, A. Hakkala, and A. Airola, “Ai in cybersecurity education-a systematic literature review of studies on cybersecurity moocs,” in *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 2020, pp. 6–10.

[5] C. A. Brayko, R. A. Houmanfar, and E. L. Ghezzi, “Organized cooperation: A behavioral perspective on volunteerism,” *Behavior and Social Issues*, vol. 25, pp. 77–98, 2016.

[6] L. Crosswell and L. Porter, “The eyes don’t lie: Addressing implicit bias in social systems,” *International Journal of Interdisciplinary Studies in Communication*, vol. 11, no. 3, pp. 1–16, 2016.

[7] D. F. Hake and R. Vukelich, “A classification and review of cooperation procedures 1,” *Journal of the experimental analysis of behavior*, vol. 18, no. 2, pp. 333–343, 1972.

[8] R. Zhang, N. J. McNeese, G. Freeman, and G. Musick, “” an ideal human” expectations of ai teammates in human-ai teaming,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, no. CSCW3, pp. 1–25, 2021.

[9] B. G. Schelble, C. Flathmann, N. J. McNeese, G. Freeman, and R. Mallick, “Let’s think together! assessing shared mental models, performance, and trust in human-agent teams,” *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, no. GROUP, pp. 1–29, 2022.

[10] A. M. Harris-Watson, L. E. Larson, N. Lauharatanahirun, L. A. DeChurch, and N. S. Contractor, “Social perception in human-ai teams: Warmth and competence predict receptivity to ai teammates,” *Computers in Human Behavior*, vol. 145, p. 107765, 2023.

[11] W. Toledo, S. J. Louis, and S. Sengupta, “Netdefense: A tower defense cybersecurity game for middle and high school students,” in *2022 IEEE Frontiers in Education Conference (FIE)*. IEEE, 2022, pp. 1–6.

[12] R. Olla, R. A. Houmanfar, S. Sengupta, E. M. Hand, and S. J. Louis, “Systematic analysis of individuals’ perspectives on cybersecurity using q methodology: Implications for research and application in behavior analysis,” *Behavior and Social Issues*, pp. 1–29, 2024.

[13] S. E. Ramlo and J. B. Nicholas, “Divergent student views of cybersecurity,” *Journal of Cybersecurity Education, Research and Practice*, vol. 2019, no. 2, p. 6, 2020.

[14] E. L. Ghezzi, R. A. Houmanfar, and L. Crosswell, “The motivative augmental effects of verbal stimuli on cooperative and conformity responding under a financially competing contingency in an analog work task,” *The Psychological Record*, vol. 70, pp. 411–431, 2020.

[15] “Training for ai based cyber security engineering (taiser).” [Online]. Available: https://github.com/sushillouis/Taiser/tree/Proto1_UI