An Investigation of Human Behaviors within Human-AI Teams using Virtual Environments

Joseph Schwalb* ‡‡, Eric Sung* ‡‡, Vineetha Menon*, Kristin Weger†,
Nathan Tenhundfeld†, Bryan Mesmer‡, Sampson Gholston‡ and Thomas Davis§

* The Department of Computer Science
† The Department of Psychology
‡ The Department of Industrial Systems Engineering and Engineering Management,
The University of Alabama, Huntsville, Alabama

§ Human Systems Integration Division, DEVCOM Data and Analysis Center, Huntsville, Alabama

Correspondence: Joseph Schwalb, Vineetha Menon; {jds0099@uah.edu, vineetha.menon@uah.edu}

***Abstract*—In this work, we investigate the adoption of Explainable Autonomous Systems within Human-AI teams and analyze the decision-making behaviors of the Human team members in a control experiment. This study is conducted in a virtual environment generated using the Unity Game Engine through the modeling of a Search and Rescue (SAR) Scenario that enables observation of player behaviors throughout each experiment. The task we assign our participants is to conduct a search for each of the 15 persons scattered throughout our virtual environment, then rescue each. To increase the complexity of our environment, we sparsely place these persons around buildings and terrain to encourage the usage of an Assistive Autonomous (AA) System such as a small drone. This work presents insight into a Human-AI team from the perspective of the Human member with a discussion of the innate difficulty of our environment in the control group of 35 participants through an analysis of proximity-based metrics. Furthermore, this work represents a gap-analysis study that is specific to our SAR scenario to understand the various dynamic facets of human-AI teaming interplay, optimization, and evaluation considerations for the effective adoption of AI-driven AA systems.**

***Index Terms*—Human-Computer Interaction, Explainable AI, Interpretable AI, Human-AI teaming, Simulation, Human Factors, Dynamic Path Planning, Target Detection**

I. Introduction

Since the European Union (EU) introduced General Data Protection Regulation (GDPR) legislature in 2016 [1], the ”right to explain” has prompted much research into the area of Explainable Artificial Intelligence (XAI) and seeded similar legislature internationally. Explainable AI aims to provide insight into what an Intelligent System observes and its ability to answer any questions any users of the system might have particularly in Human-Computer Interaction (HCI) applications. [2].

A key issue in this nascent field is a lack of common terminologies with widely shared definitions [3], enabling confusion between terms that have different meanings. Work in [2], [4]–[7] narrow the definitions of commonly misused terms such as ”Explainability” and ”Interpretability”. According to [4], [8], Explainability is the cognitive workload required to understand the details and reasons a model provides to make its functioning clear. To this extent, useful information to share with a user can include a model's internal representation of learned classes through model-specific or model-agnostic methods. For example, an explainable system might utilize a Gradient Class Activation Map (G-CAM), which leverages the gradients found during backpropagation to extract information about the learned features. Unlike Explability, Interpretability describes the transparency a model can provide by breaking down this characteristic into three categories: Simulatability, Decomposability, and Algorithmic Transparency. Where Simulatability describes the ability of a model to be thought about strictly by a human (simulated). Decomposability describes the ability to explain each part of the model (data, model, and prediction). Finally, Algorithmic Transparency describes the ability of the user to understand the process a model employs to make a prediction.

Importantly, the example given is a subjective measurement of interpretability, and a recent effort to measure AI Interpretability relies on behavioral analysis, as well [9], [10]. These behaviors, by human nature, are not objective behaviors as they can be influenced by the physical state, personal/emotional factors, or life experiences of the person [11]. To address this shortcoming, we propose describing Interpretability as the observed deviation from the ideal interaction with an AI Agent.

The SAR environment presented in our prior studies [12], [13], models an airborne AI-based target detection (there hostage detection, here person detection) system which is capable of both maneuvering challenging terrains and locating people in the environment. The high maneuverability of drones with onboard cameras makes them an ideal candidate for the

These authors contributed equally to this work. ‡‡

This work was supported by the Research Development and Engineering Command (DOD - Army - AMC) [grant number W911NF2220001]

surveillance of hazardous areas under high-risk situations, such as SAR situations, battlefields, or disaster-relief scenarios. The conventional use of drones often relies on human operators' manual controls, but an autonomous navigation system for an assistive drone can eliminate the controlling task. In SAR missions, the drone can provide an assistive role as an AI agent that is capable of automated search for targets of interest (people or other objectives) in complex operational environments with varying degrees of autonomy. Thus, AA integrated with AI (AI/AA) systems is poised to provide real value by reducing the cognitive load required for a human operator to deploy the system successfully. This AI/AA system can also improve the mission success rate by supporting the search function of this particular task through the sharing of information, leading to an educated decision-making process for the player. Therefore, in this study, we extend our prior research in a controlled SAR scenario to observe the decision-making behaviors of a player (human participant) as they search our virtual environment for persons to be rescued without access to the drone-based AI/AA capability in this experiment. This allows comprehensive performance analysis considerations to realize an integrated human-AI teaming technology. It also signifies the role of AI/AA systems as critical assistive decision support systems to navigate challenging environments and accomplish complex tasks such as in a SAR situation.

II. Methods

A. SAR Simulation Environment

We designed the SAR virtual environment using the Unity Engine [14]. We combined common off-the-shelf assets with our own custom designs to emulate a SAR scenario, complete with complex terrain and buildings to make exhaustive searching a non-trivial task. The goal of this SAR mission is to identify and rescue all persons, 15 of which are sparsely placed throughout the environment. We integrated a drone-based XAI/AA system in this environment and AI-based target detection capabilities to automatically identify persons and mark locations in the form of a red dot inside a shared minimap to assist the human players for a targeted rescue goal as discussed in detail in our previous work [12], [13]. However, in our previous work [12], [13], the human players had access to control and interact with the drone-based AI/AA system. However, in this control experiment, we conduct a gap-analysis study to further analyze and derive human player behaviors heuristics discussed in this section. This study provides the XAI context required to understand the variations in a human player's decision-making behaviors with respect to the proximity of objectives and the difficulty characterized by this specific environment. In particular, we are interested in understanding the role of XAI in AI/AA systems for providing the reasoning behind their decisions, improving its trustworthiness to the human user, and exhibiting subsequent influence on human-decision making for a feasible adoption of AA/AI systems in our daily lives.

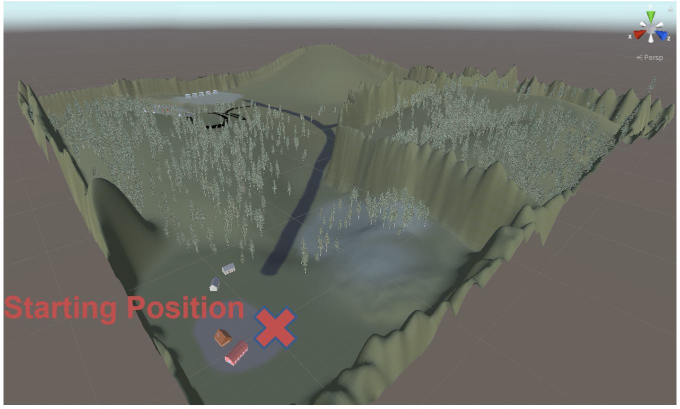

Fig. 1. Aerial terrain view of the SAR experiment environment. Participants begin at the starting position, marked with the red ”X”.

B. Data Collection Procedure

The data collection process for this experiment was conducted by acquiring 35 participants from a pool of undergraduate university students. First, informed consent was received from each participant. Next, we conducted a survey to assess how comfortable our participants are with using automation. Then, the participants were informed with detailed instructions on how to move and interact within the environment. At this moment, the participants were presented with the opportunity to ask questions before beginning. After the experimental procedure was explained, the participants began performing the informed task of rescuing every person (also defined as ’objectives’ in this experiment) observed in the simulation environment. Once, the task was completed, individual participants' performance data were recorded.

For human player behavior evaluation purposes, in this study, we tracked the identifiers of rescued ’objectives’ and the positions of the participants throughout the duration of the experiment. The positions were each recorded in a 3D coordinate system (x, y, z) in meters. The positions of both the objectives and the participants were each converted to a graph node with the geographical position and its nearest connected neighbor information. We use this information to generate the ordering of captured objectives as well as the ordering most observed to visualize common behaviors across the SAR simulation environment. The presentation of this visualization allows us to understand and evaluate the interplay of human player decision-making aspects to an optimal AI-based path planning solution discussed in the following nearest objective search section. The goal of this paper is to identify the various human decision-making behavioral patterns observed in the SAR scenario in order to realign it with the expected optimal AI-based path identified for SAR mission success. The key is to encourage and empower the human players to confidently utilize the XAI/AA systems for improved transparency in AI decisions, human-AI trust, and mission success goals.

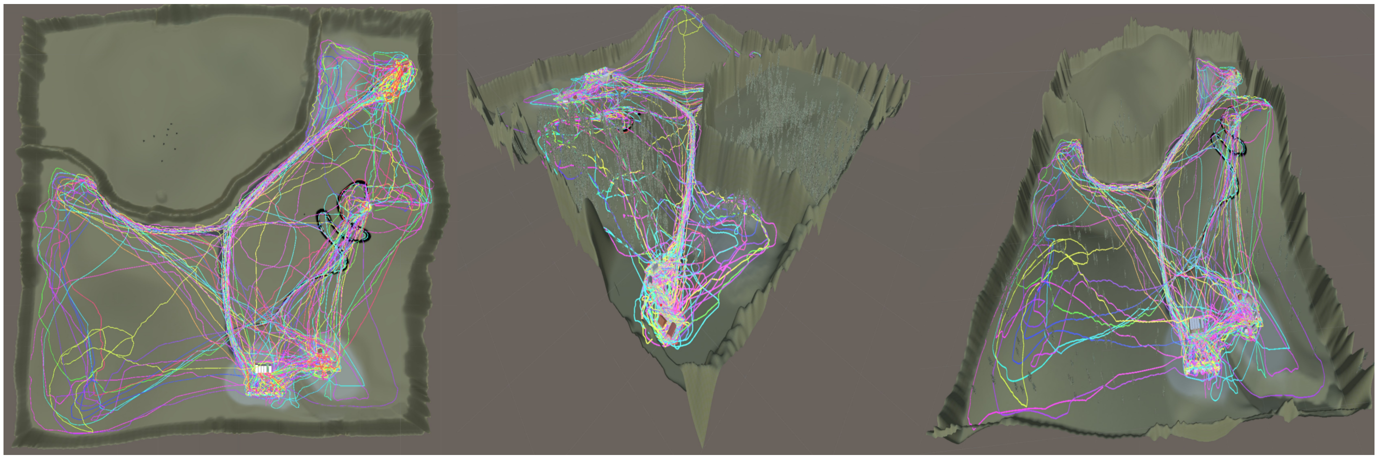

Fig. 2. Player behaviors throughout the duration of the experiment for all participants. From left to right, the first image depicts a top-down perspective of the entire simulation environment. The second image changes perspective to the starting location, and the third image shifts perspective to a building complex located on the opposite side of the starting location.

C. Nearest Objective

It is essential to outline guidelines for quantifying the degree of alignment (decision-making performance) that an AI/AA system can attain in a Human-AI teaming environment. Ultimately, the ideal path to complete the SAR task with minimal duration can be generated by AI path planning search algorithms such as Djikstra's [15] or A* [16]. However, without knowledge of the world, the human versus AI behaviors exhibited will not play out as precisely as expected. For this reason, we want to measure the change in fixed (globally in the environment) as well as relative player proximity distance to each objective throughout the duration of this experiment.

For this, we formulated a unique objective measurement strategy and heuristic to study the human-AI teaming interactions for the SAR mission. Here we represent each objective position (as fixed) and the player position (variable) nodes in our objective proximity graph. This is accomplished by first employing a proximity search algorithm, specifically, the Nearest Neighbor Search (NNS), to determine the point of reference (player position) node to compute their proximity distance to the nearest fixed objectives in our environment. In general, NNS is a simple search algorithm that finds the closest node c to the given node g_n from a list of available objective nodes N. Euclidean distance was used to compute the node distances. In our case, we are interested in the evolution of player paths followed to rescue $N = 15$ objectives with the progression of time throughout the experiment. Note that the size of N decreases with respect to time as the evolutionary players rescue the objectives to complete their SAR task. See Algorithm 1 for the pseudo-code of the NNS variation used to compute and construct the player to objective proximity map.

The NNS algorithm described in 1 was applied at each time step to generate the closest nodes throughout the duration of the experiment. Fig. 3 illustrates the plot of the optimal path as constructed using the Nearest Neighbor Heuristic as a solution to the Travelling Salesman Problem [17] (expected behavior in blue) versus the actual human player behavior observed (green). Fig. 2 describes the player behaviors and paths taken to rescue all the hostages in the SAR environment.

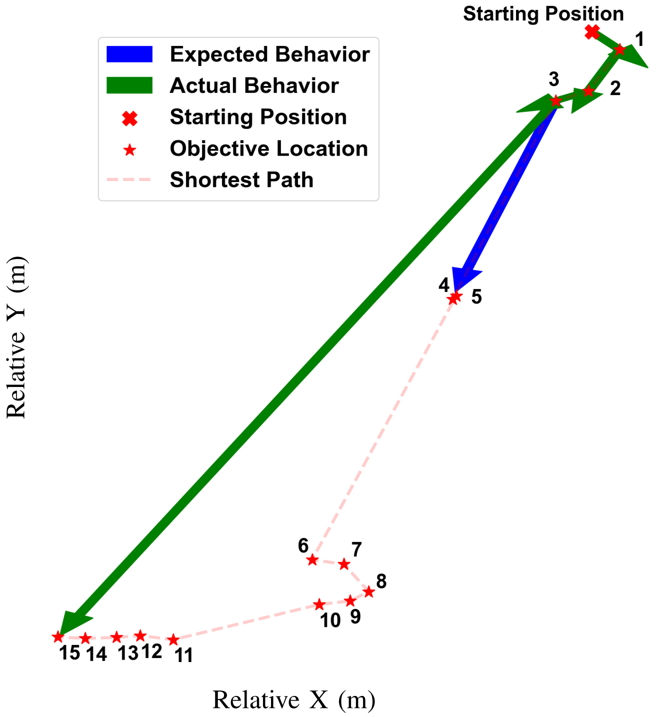

Fig. 3. A comparison of our hypothesized behavior and the actual behavior exhibited most often for all experiments in the context of the starting location, objective locations, and the shortest possible path.

D. Measuring Task Difficulty

Understanding the SAR task completion difficulty that our first round of participants experienced during experimentation and their player behaviors as its result is key to making conclusions about the efficacy of AI/AA support systems that

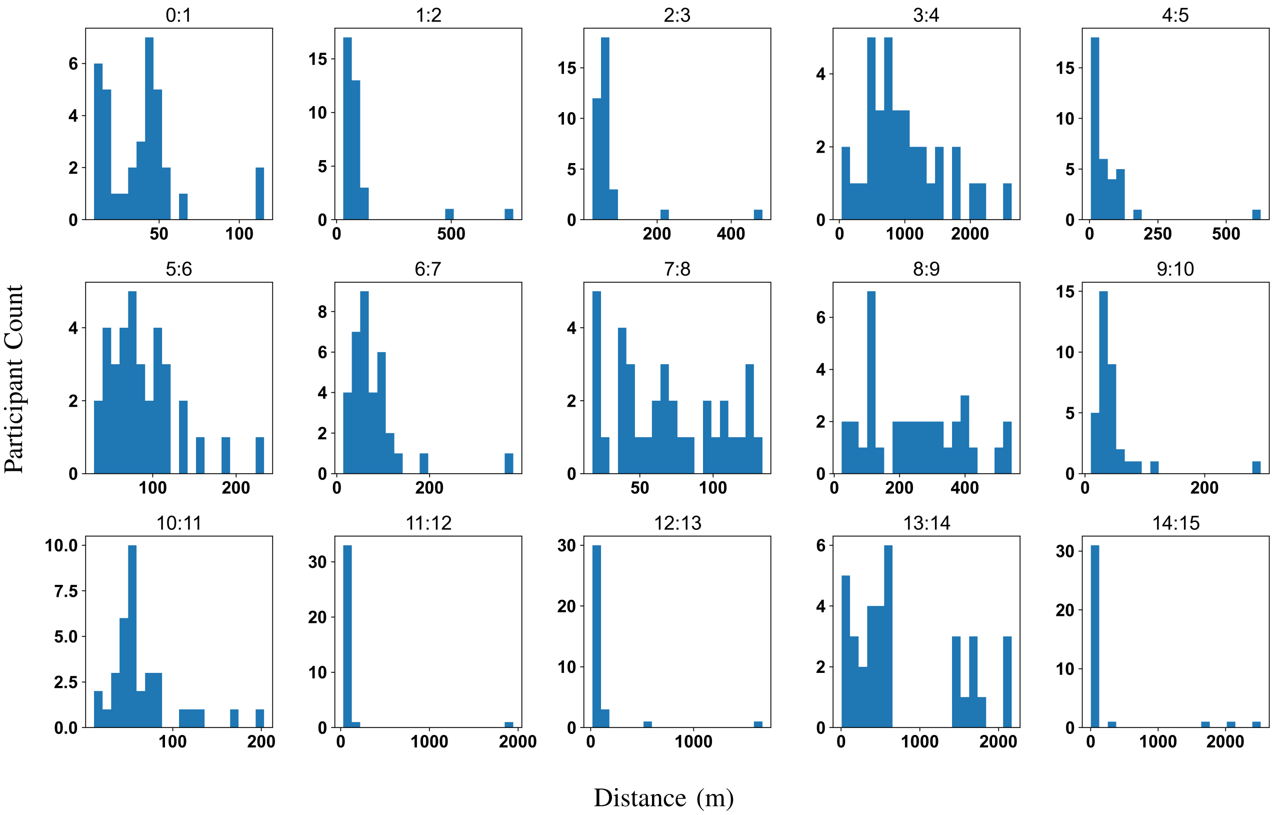

Fig. 4. The Euclidean distance to the next objective in the relative ordering exhibited by each participant for all experiments.

Algorithm 1 Nearest Neighbor Search Algorithm

```
$g_n \leftarrow$ given player position $(g)$ as a node
$N \leftarrow$ list of all unrescued objective nodes
$c \leftarrow$ None    // closest objective node
$dist_c \leftarrow \inf$
for $n \in N$ do
    $dist_n \leftarrow dist(g_n, n)$
    if $dist_n < dist_c$ then
        $c \leftarrow n$
        $dist_c \leftarrow dist_n$
    end if
end for
return $c$
```

have advanced features such as AI Explanations, Interpretable Models, or Human-Feedback Reinforcement Learning (HFRL) [18]. Since we scattered the objectives or persons throughout the SAR environment in such a way as to encourage exploration of the entire map, we want to understand the implicit impact of objective positioning on the outcome of the experiment. Using the Euclidean distance metric in the time-evolutionary NNS algorithm outlined previously in 1 provides a coarser view of performance for all objectives, which may occlude any complexity encountered through the search conducted for a particular objective or set of objectives. To measure this, we observed the time spent and distance traveled by all participants to rescue each person. We computed the proportion of search costs associated with individual search and rescue instances and identified two of the most difficult objectives in terms of the highest relative search costs. Figs. 2 captures the player path navigation information overlaid on the complex terrain view. Fig. 4 signifies the corresponding relative difficulty in objective rescue task completion as dictated by the challenging terrains in our SAR environment.

III. Results

From this SAR control experiment, we observed that the behavior exhibited by our participants did not align with our expected objective rescue ordering. We found that there was no uniformity in the order in which the participants rescued persons. Instead of adhering closely to the AI-computed optimal path through our NNS algorithm, we observe a distinct human decision-making deviation behavior that changes the ordering of rescued persons. In Fig. 3, we describe this unexpected behavior in the context of the expected behavior, objective locations, and the optimal path. While our expected objective ordering is consistent with the shortest path (1-15, in ascending order), the objective ordering most exhibited was [1, 2, 3, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 4, 5]. Considering the decision depicted in Fig. 3, this ordering highlights the exploratory nature of our participants to accomplish this task.

The difficulty in searching for each objective, in a relative ordering, is described in Fig. 4. Each subplot enumerates the distance traveled to rescue from person i to the person $i + 1$, as denoted in each subplot title in the form ”$i : i + 1$”. The special case of ”0:1” details the distance traveled from the starting point to rescue the first person. We find the most difficult search was experienced between persons 3:4 and 13:14, with a mean distance traveled of 985.75m and 776.52m respectively. Table I presents a detailed summary of these costly inter-objective searches and the experiment summary. Since the mean distance traveled for each experiment was 3067.59m, the cost of the search & rescue for these persons consumes approximately 57% of the entire distance traveled for each experiment. From Table I, it can be noted that we observe a mean time of 847.79s to complete this experiment, which can be equated to an average speed of 3.62m/s. Finally, the proximity calculations yield a mean distance of 100.56m with a standard deviation of 147.15m. Fig. 5 describes the time taken versus the distance traveled on complex terrains to rescue all the objectives or persons in this experiment for all the participants. It emphasizes the sparse distribution of objectives throughout the challenging SAR terrains. Herein, the adoption of AI/AA systems technology would be of great assistance in exploration-tasks-oriented human-AI teaming.

Therefore, our future research directions in SAR mission situations will explore the impact different methods of Explainable AI and Interpretable AI have on the distributions presented in Fig. 4. With this in mind, a good teaming result would manifest as a mean value closer to zero, a smaller standard deviation, or a mix of both. Whereas a no-impact or bad teaming result would manifest as a greater mean value, a larger standard deviation, or both.

TABLE I
STATISTICAL SUMMARY OF THE HUMAN-AI TEAMING EVALUATION AND INTER-OBJECTIVE SEARCH STATISTICS

Units	**Distance (m)**				**Time (s)**
	Inter-Objective		**Experiment**		
Statistics	***3:4***	***13:14***	***NN Proximity***	***Total***	***Time***
mean	985.75	776.52	100.56	3067.59	847.79
std	593.04	689.68	147.15	967.73	368.68
min	33.84	7.15	< 1	1291.42	335.58
25%	571.77	286.34	8.72	2289.64	631.38
50%	821.13	527.10	30.07	3082.67	783.98
75%	1323.35	1427.37	130.22	3780.32	1007.58
max	2630.96	2165.74	740.98	5083.17	1834.38

IV. CONCLUSION

Through previous studies, we identified the need for a more detailed understanding of player behaviors without the usage of an XAI/AA system. The goal of this work is to highlight the need for a more inclusive interdisciplinary study from the systems, computer science, and psychological perspectives to understand the design, evaluation, validation, and feasibility of XAI/AA systems technology in Human-AI teaming and interactions for SAR missions. In this study, we found that player behaviors tended to stray from the shortest path as found using an AI path optimization algorithm, and this decision directly contributed to the complexity of our environment. Through this finding, we calculated a graph distribution of distances to the nearest objectives for all experiments, empirically associating our observed human behavior through statistical analysis. Therefore, this work presents further insights into the design considerations for effective human-AI teaming in the context of challenging terrains and the adoption of AI/AA systems and player proximity to objectives as some of the driving behaviors that influence human decision-making versus AI optimal path decisions.

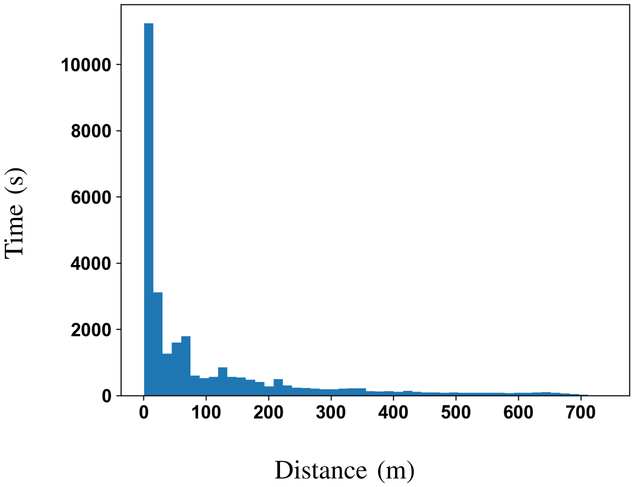

Fig. 5. The distance to the nearest objective, calculated using the nearest neighbor algorithm using all participant data.

V. FUTURE WORK

This work details the complexity of our SAR environment for Human-AI teaming through the analysis of Human behaviors in a controlled study. Future work will explore novel XAI features to provide an evaluation of the explanation effectiveness of the AI model and its influence on the frequency and duration of Human-AI interactions. Further, the recent advancements in HFRL [18] and the alignment of AI systems [19], [20] have interesting applications in this area to enhance the capabilities of an AI/AA system through a human-in-loop feedback framework for a dynamic teaming aspect that changes over time.

ACKNOWLEDGMENT

This research was supported by the Army - Subcontract through Northeastern University entitled ”Testing & Evaluation for Soldier-Device Teaming Compatibility, Vulnerability, and Durability in Emergent Situations: How to assess the efficacy of AI in human agent teams” through W911NF2220001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the

Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] "General Data Protection Regulation (GDPR) – Official Legal Text," Apr. 2016.

[2] M. Bellucci, N. Delestre, N. Malandain, and C. Zanni-Merk, "Towards a terminology for a fully contextualized XAI," *Procedia Computer Science*, vol. 192, pp. 241–250, Jan. 2021.

[3] T. Miller, "Explanation in Artificial Intelligence: Insights from the Social Sciences," Aug. 2018. arXiv:1706.07269 [cs].

[4] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, June 2020.

[5] A. Rawal, J. McCoy, D. Rawat, B. Sadler, and R. Amant, "Recent Advances in Trustworthy Explainable Artificial Intelligence: Status, Challenges and Perspectives," Nov. 2021.

[6] F. Doshi-Velez and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," Mar. 2017. arXiv:1702.08608 [cs, stat].

[7] V. Beaudouin, I. Bloch, D. Bounie, S. Clémençon, F. d'Alché Buc, J. Eagan, W. Maxwell, P. Mozharovskyi, and J. Parekh, "Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach," *SSRN Electronic Journal*, 2020.

[8] Z. C. Lipton, "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery.," *Queue*, vol. 16, pp. 31–57, June 2018.

[9] A.-M. Nussberger, L. Luo, L. E. Celis, and M. J. Crockett, "Public attitudes value interpretability but prioritize accuracy in Artificial Intelligence," *Nature Communications*, vol. 13, p. 5821, Oct. 2022. Number: 1 Publisher: Nature Publishing Group.

[10] S. R. Hong, J. Hullman, and E. Bertini, "Human Factors in Model Interpretability: Industry Practices, Challenges, and Needs," *Proceedings of the ACM on Human-Computer Interaction*, vol. 4, pp. 1–26, May 2020. arXiv:2004.11440 [cs].

[11] "What factors can affect behaviour? - Principles for effective support," Jan. 2020.

[12] J. Schwalb, V. Menon, N. Tenhundfeld, K. Weger, B. Mesmer, and S. Gholston, "A Study of Drone-based AI for Enhanced Human-AI Trust and Informed Decision Making in Human-AI Interactive Virtual Environments," in *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, pp. 1–6, Nov. 2022.

[13] D. Pham, V. Menon, N. Tenhundfeld, K. Weger, B. Mesmer, S. Gholston, and T. Davis, "A Case Study of Human-AI Interactions Using Transparent AI-Driven Autonomous Systems for Improved Human-AI Trust Factors," in *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, pp. 1–6, Nov. 2022.

[14] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, and D. Lange, "Unity: A General Platform for Intelligent Agents," May 2020. arXiv:1809.02627 [cs, stat].

[15] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959. Publisher: Springer.

[16] P. E. Hart, N. J. Nilsson, and B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, pp. 100–107, July 1968. Conference Name: IEEE Transactions on Systems Science and Cybernetics.

[17] M. M. Flood, "The Traveling-Salesman Problem," *Operations Research*, vol. 4, no. 1, pp. 61–75, 1956.

[18] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, and B. Ge, "Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models," Apr. 2023. arXiv:2304.01852 [cs].

[19] OpenAI, "How should AI systems behave, and who should decide?." https://openai.com/blog/how-should-ai-systems-behave. (Accessed: 2023-04-12).

[20] OpenAI, "Our approach to alignment research." https://openai.com/blog/our-approach-to-alignment-research. (Accessed: 2023-04-12).