Measuring State Utilization During Decision Making in Human-Robot Teams

Saurav Singh
ss3337@rit.edu
Rochester Institute of Technology
Rochester, New York, USA

Jamison Heard
jrheee@rit.edu
Rochester Institute of Technology
Rochester, New York, USA

ABSTRACT

Efficient team design necessitates a comprehensive understanding of human factors, encompassing abilities, limitations, and internal states. In human-robot teaming research, recent efforts explore integrating emotions, workload, fatigue, and stress into decision-making using deep reinforcement learning. Despite promising results, the black-box nature of these algorithms raises questions about the consistent reliance on human internal states or their consideration as information or noise in the decision-making process. This study introduces a state utilization (SU) metric to measure the reliance of reinforcement-based agents on each state feature. This metric is validated on data from the Cartpole environment by OpenAI and a human-robot teaming experiment using NASA MATB-II environment. The SU provides insight into the relevance and usage of state features and human data modalities by the robot, showing clear trends based on the nature of the tasks and offering an understanding of why the RL agent takes certain actions. This, in turn, enhances the explainability of the RL agent's policy used for human robot teaming.

CCS CONCEPTS

• **Human-centered computing** → *Usability testing*; **Scientific visualization**; *Laboratory experiments.*

KEYWORDS

Reinforcement Learning, Human-Robot Teaming, State Utilization, Decision Making, Physiological Data, Workload

ACM Reference Format:
Saurav Singh and Jamison Heard. 2024. Measuring State Utilization During Decision Making in Human-Robot Teams. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3610978.3640676

1 INTRODUCTION

Human's complex and unpredictable behaviors significantly influence human-robot teaming dynamics, emphasizing the need to comprehend individual agent's abilities, adaptation, and decision-making processes. Successful human-robot team design requires an intimate understanding of these dynamics, including external states (such as position, velocity, head pose and gaze) and internal states (such as emotions, workload, fatigue and stress) [15].

The human body is a sophisticated, self-adaptive system that regulates internal states to respond to environmental factors, with observable physiological signal changes. Some examples are human's pupil dilating in response to the emotions like fear [13], an increase in heart rate variability can indicate high levels of stress [12], or heart rate and respiration rate are sensitive to cognitive workload [11]. These physiological measurements provide an indirect measure of various human performance constructs, such as fatigue [16], stress [8], and workload [9, 10]. Thus, a robot teammate can greatly benefit from knowing the human teammate's emotional and physical states [7][14][2][17], similar to how human-human teams operate [5].

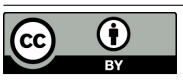
The intricate relationship between human internal states and human-robot team performance, coupled with the inherent unpredictability of human behavior [3], creates uncertainties in how robots utilize human data in decision-making. This challenge is heightened in reinforcement-learning-based algorithms, known for their limited transparency and explainability [22], impacting human trust and overall team dynamics [18].

Addressing this research gap, our study builds on a previous modality utilization metric [21], extending it to reinforcement learning. This work introduce the State Utilization (SU) metric that quantifies the utilization of state features and human data modalities by the robot, thus highlighting their importance and contributing to improved explainability of the RL agent's policy. The SU metric was evaluated in the Cartpole environment by OpenAI gym, followed by ablation studies. Applying this metric to data from a human-robot teaming experiment on the NASA MATB-II [19][20], this study identifies distinct trends that aligned with task characteristics.

This study's contribution lies in quantifying the reliance of decision networks on specific modalities, emphasizing their crucial role in shaping the RL agent's behavior. This insight paves the way for refined decision-making mechanisms, enhancing overall performance across diverse tasks and environments, and addressing the critical need for transparency and trust in human-robot teaming dynamics.

2 METHOD

The State Utilization (SU) metric is an extension of the Modality Utilization (MU) metric [21], which was inspired by permutation feature importance [1][6]. Given an RL decision network (such as a Q network for Q-learning algorithm, or an actor network for an actor-critic algorithm) F_θ and $\mathcal{D}_{su}$ (a subset of replay memory $\mathcal{D}_{reply}$) with M state features. State utilization SU_i is computed by breaking the association between the input state feature S_{ni} and the network output Y and calculating the resulting difference in output



HRI '24 Companion, March 11–14, 2024, Boulder, CO, USA

ACM ISBN 979-8-4007-0323-2/24/03
https://doi.org/10.1145/3610978.3640676

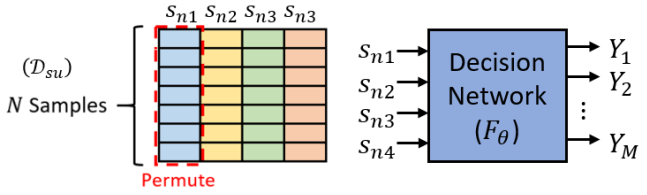


Figure 1: Example of permuting state feature samples S_{ni} to break the association between the input state feature S_{ni} and the decision network output Y.

Y from the original batch data $\mathcal{D}_{su}$. SU metric is more accurate with more samples, i.e., ideally $\mathcal{D}_{su}$ should be the same as $\mathcal{D}_{reply}$.

The association between a state feature S_{ni} and the output Y is broken by permuting/shuffling the corresponding state feature (S_{ni}) randomly amongst the samples, while keeping the remaining state feature (S_{nj}, $j \neq i$) the same, as shown in Figure 1. Let independent samples from the replay buffer in $\mathcal{D}_{su}$ be of the form $\mathcal{D}_{su} = (Y, S_{n1}, S_{n2}, \cdots, S_{nM})$ be

$$Sample^{(a)} = (Y^{(a)}, S_{n1}^{(a)}, S_{n2}^{(a)}, \cdots, S_{nM}^{(a)})$$

$$Sample^{(b)} = (Y^{(b)}, S_{n1}^{(b)}, S_{n2}^{(b)}, \cdots, S_{nM}^{(b)})$$

A new permuted batch $\mathcal{D}_i$ is then generated, where samples of the i^{th} state feature in $\mathcal{D}_{su}$ are permuted ($S_{ni}^{(a)}, S_{ni}^{(b)}$) as

$$Sample^{(b)}_{permuted,i} = (Y^{(b)}, S_{n1}^{(b)}, S_{n2}^{(b)}, \cdots, S_{ni}^{(a)}, \cdots, S_{nM}^{b}) \quad (1)$$

Let the output of the RL decision network F_θ be Y during inference with the original batch data $\mathcal{D}_{su}$ and Y_i during inference be with the permuted batch data $\mathcal{D}_i$, where samples of ith state feature ($S_{ni}^{(a)}, S_{ni}^{(b)}$) are permuted:

$$Y = F_\theta(Sample^{(b)}) \quad (2)$$

$$Y_i = F_\theta(Sample^{(b)}_{permuted,i}) \quad (3)$$

The state utilization of the i^{th} state feature (SU_i) for the RL decision network F_θ can be computed by observing the change in the model output Y during inference with the original batch data $\mathcal{D}_{su}$ and the permuted batch data $\mathcal{D}_i$. Observing the euclidean distance between Y and Y_i for Q-learning algorithms or KL divergence for actor-critic algorithms, a reduced discrepancy suggests that the rearrangement of samples related to the i^{th} state feature minimally affects the decision network's output Y, indicating limited utilization of the state feature. Conversely, an increased discrepancy implies that shuffling the i^{th} state feature samples significantly influences the decision network's output Y, signifying a substantial utilization of the state feature. State utilization of the i^{th} state feature (SU_i) is then defined as:

$$SU_i = \frac{\|Y_i - Y\|}{\sum_{j=1}^{S} \|Y_i - Y\|} \quad (4)$$

Algorithm 1: State Utilization for RL

Initialize the RL decision network F_θ, learned model parameters θ, replay memory $\mathcal{D}_{replay}$;
Sample a batch of data $\mathcal{D}_{su}$ from replay memory $\mathcal{D}_{replay}$;
Compute decision network output Y, Eq. 2;
for *each state feature* s_{ni} **do**
  Randomly permute the samples of state feature s_{ni} while keeping the state features s_{nj}, $j \neq i$ unchanged;
  Compute decision network output Y_i with permuted state feature s_{ni}, Eq. 3;
end
for *each state feature* s_i **do**
  Compute State Utilization (SU_i) using $SU_i = \frac{\|Y_i - Y\|}{\sum_{j=1}^{S} \|Y_j - Y\|}$, Eq. 4;
end

3 ABLATION STUDIES ON CARTPOLE

SU metric was validated on OpenAI Gym's Cartpole environment, a standard benchmark for reinforcing learning algorithms. The setup involves a cart moving horizontally with an attached pole, aiming to balance it. The system state includes cart position (S_{n0}), velocity (S_{n1}), pole angular position (S_{n2}), and pole angular velocity (S_{n3}). The agent can take two actions: apply a force to move the cart left or right. Episodes conclude if the pole exceeds a specific angle or the cart moves beyond a set range. Successful solving is maintaining an average reward of 195 or higher over a continuous 100-episode period.

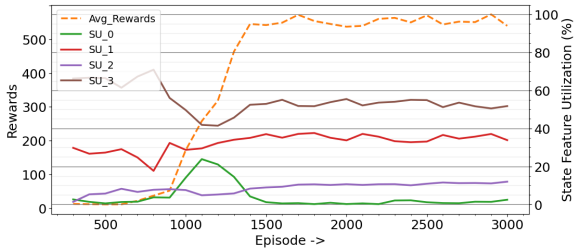


Figure 2: Average Rewards and State Utilization (SU) across episodes for the CartPole environment.

A Double Deep Q-Network (DDQN) successfully solved the Cartpole environment. Figure 2 shows average rewards and state utilization (computed using algorithm 1). $\mathcal{D}i$ had 1280 samples, ten times the batch size of 128. The SU metric, assessed every 100 episodes, highlights angular velocity ($Sn3$) as the most utilized state feature, while cart position (S_{n0}) is the least utilized. Results indicate a dynamic shift in feature importance over time. Initially, S_{n0} had over 20% utilization, dropping to nearly 0% after episode 1600, suggesting redundant information.

The optimized policy disregards S_{n0} entirely, and training the RL agent without it led to expedited performance improvements, as depicted in Figure 3. This highlights the potential for leveraging a simplified state space for easier exploration. This raises questions about the importance of information in individual state features and

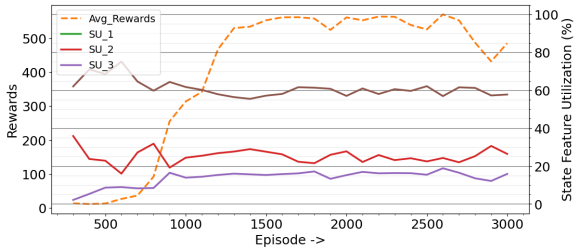


Figure 3: Average Rewards and State Utilization (SU) across episodes for the CartPole environment without S_{n0}.

the extent to which agents rely on redundant or noisy data, insights provided by the proposed state utilization metric. Additionally, the agent was trained with an extra state feature, random uniform distribution noise in the range $[-1, 1]$.

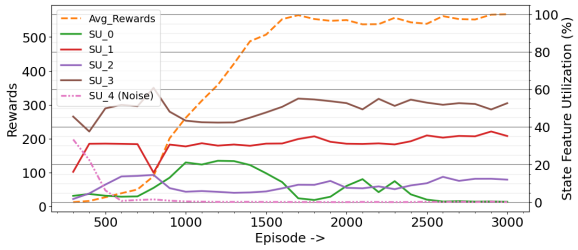


Figure 4: Average Rewards and State Utilization (SU) across episodes for the CartPole environment with a random noise state feature (S_{n4}).

In Figure 4, the RL agent quickly learned to ignore the noisy state feature. However, solving the environment took longer due to increased complexity, making the state space more challenging to explore. The state utilization metric reveals the impact of redundant and noisy information in the RL agent's state space, offering insights into the explainability aspect of reinforcement learning.

4 MEASURING UTILIZATION OF HUMAN DATA IN HUMAN ROBOT TEAMING

Human-robot collaboration is essential for maximizing productivity, as robots excel in speed, precision, and hazardous tasks, complementing human creativity and adaptability to enhance overall efficiency. Recognizing teammates' mental and emotional states in teamwork improves collaboration and fluency, while acknowledging fatigue or workload anticipates potential performance decline. However, this internal state information can be extremely noisy. Thus, it is essential to understand if RL-based agents leverage this information or learns to ignore it using the state utilization metric. The potential insights of the SU metric are demonstrated through analyzing the utilization of human data in a previous human-robot teaming study [19][20].

4.1 Summary of previous study

The paper [19][20] introduces a human-aware decision-making paradigm for enhancing human-robot collaboration in high-stress scenarios using reinforcement learning (RL). It aims to adapt a robot's interactions based on human workload states, leveraging the NASA Multi-Attribute Task Battery (MATB) environment [4] (Figure 5) to simulate real-world challenges. Participants engage in four concurrent tasks: Tracking, System Monitoring, Resource Management, and Communications, representing scenarios like target tracking, system parameter monitoring, resource management, and response to audio commands.

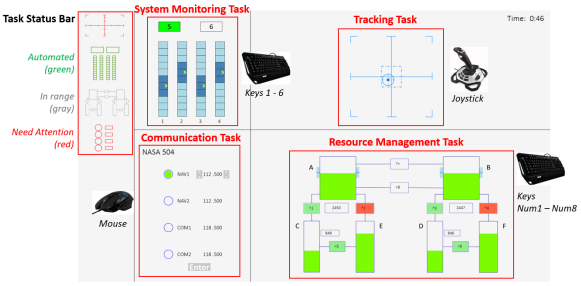


Figure 5: The NASA Multi-Attribute Task Battery-II Env.

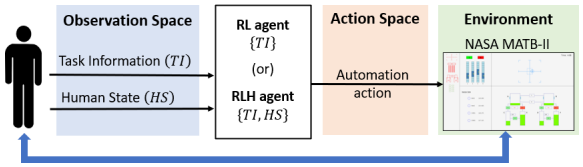


Figure 6: Adaptive Human-Robot Teaming architecture with human state estimates augmented to *RL*'s observation space.

Nine participants undergo a 15-minute training, followed by a 52.5-minute baseline trial with a rule-based adaptive scheme. Workload conditions are manipulated to create scenarios of underload (UL), normal load (NL), and overload (OL). Participants then experience experimental trials with *RL* or *RLH* agents, guided by a Soft Actor-Critic (SAC) agent making automation decisions in two state spaces. The first relies on task interaction data (*RL*), while the second augments it with estimated human workload states (*RLH*), as illustrated in Figure 6. Physiological, workload, and task-related data are collected for assessment.

Results show that *RL* achieves the highest rewards but also the highest workload, while RB has lower rewards and the lowest workload. *RLH* maintains a lower workload but achieves the lowest rewards, excelling in overload conditions. *RL* outperforms in system monitoring and resource management, while *RLH* excels in tracking and communication. Automation time analysis revealed *RB* focused on automating resource management, *RL* on system monitoring, and *RLH* on communication. Despite *RLH* reducing perceived workload, the more complex state space may have hindered reward achievement. Despite reducing perceived workload, *RLH* faces challenges due to a more complex state space. In conclusion, the paper underscores the potential of human-aware reinforcement learning to revolutionize team collaborations and enhance overall performance in dynamic and high-stakes task environments.

4.2 Measuring human data utilization

The proposed SU method for reinforcement learning was used to observe the utilization of human data in the study described in Section 4.1. SU metrics were computed using SAC agents' trained actor network and (state, action, next state) pairs from the prior investigation. Due to SAC agents optimizing a stochastic policy, KL divergence between actor network output probability distributions (Eq. 4) was used instead of euclidean distance. Figures 7 and 8 display state utilization for RL and RLH agents trained on data from trial 1 (*RB*), trial 2 (*RL*), and trial 3 (*RLH*).

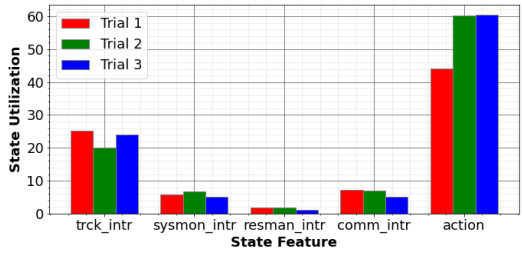


Figure 7: State Utilization for the RL agent trained on task interaction data in the NASA MATB-II experiment.

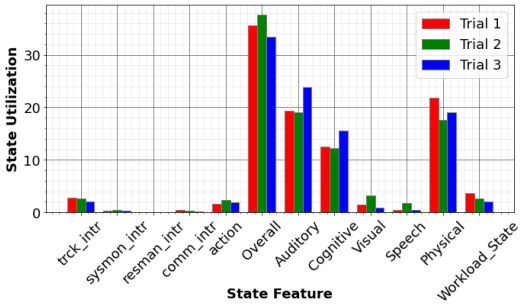


Figure 8: State Utilization for the RLH agent trained on task interaction and human internal states data in the NASA MATB-II experiment.

Figure 7 illustrate that for the *RL* agent, the agent relied the most on the last task automated (*action*), followed by tracking task interaction data (*trck_intr*). The *RLH* agent had access to the estimated human workload data and the SU metric reveals that the agent relies the most on the overall workload estimates, followed by the auditory, physical, and cognitive workload components, as shown in Figure 8. The trends were similar with the agents trained on data collected during trial 1 (*RB*), trial 2 (*RL*), and trial 3 (*RLH*).

5 DISCUSSION & CONCLUSION

This study introduced and validated the State Utilization (SU) metric, assessing an RL agent's reliance on individual state features. Preliminary validation in a cartpole environment demonstrated the agent's ability to ignore noisy/non-informative states. Applied to a human-robot teaming scenario with human states, the SU metric revealed higher reliance on human workload data, guiding more automation decisions without categorizing human data as noise. Additionally, distinct trends in reliance on overall workload, physical, auditory, cognitive workload features emerged, providing insights into the rationale behind specific RL agent actions.

Utilizing overall workload the most to determine automation decisions (or no-automation) is how typical rule-based adaptive autonomy agents are designed; thus, it is interesting that the *RLH* agent had similar reliance on overall workload despite relatively more information being available. However, the agent did utilize other workload components, which may have promoted more effective decisions. For example, auditory workload was the second most utilized state and is also only present during the communications task. This task was also the most difficult task for participants to complete and the most common task for the agent to automate. Similarly, the continuous fine-motor control required for the tracking task may be why physical workload was the next highest utilized state. There are a few discrepancies though, as speech and visual workload were the lowest utilized workload states. This may be attributed to redundant information. For example, speech and auditory workload were only associated with the communications task and were not required for any other task. Thus, the RLH agent may rely on a single state to gain some understanding of the task due to redundancy. Auditory workload may have been chosen for this state, as speech was only required part of the time (e.g., the task was being automated or the communications request was directed at a different aircraft), but auditory processing was always required. A similar case may be made for visual workload, as cognitive workload was utilized much more than visual and all tasks required both of these components.

The experimental design maintains a task-agnostic observation space for *RL* and *RLH* agents, and the state utilization may differ with task-specific features. Future studies will focus on leveraging the SU metric to encourage AI reliance on multiple modalities, instead of just human data during training. While preliminary results are promising, further investigation into the effects of the replay buffer on SU metric is necessary. The use of a replay buffer introduces potential recency bias, evaluating the metric predominantly on more recent samples. Additionally, the SU metric accuracy may be affected if the RL agent utilizes a sparse state space. Despite these considerations, the SU metric shows potential in guiding the underlying reliance of a decision network on specific modalities during training, and future studies aim to develop state utilization-based training for reinforcement learning.

The State Utilization (SU) metric is a groundbreaking advance in Human-Robot Interaction (HRI) research, quantifying RL agents' reliance on specific state features, including human data in human robot teaming scenarios. It streamlines RL system designs by focusing on essential information in the state space, enabling faster agent training, particularly in scenarios where human interaction data collection is costly. It also significantly enhances RL agent's policy explainability, paving the way for a transformative future. Future emphasis on AI's reliance on multiple modalities via SU-based training promises to revolutionize decision-making, elevating overall performance across diverse tasks. The potential for context-based training, with specific modifications, enables personalized models using rich human metadata, positioning the SU metric as a valuable tool in advancing HRI research and RL training methodologies.

REFERENCES

[1] Leo Breiman. 2001. Random forests. *Machine Learn.* 45 (2001), 5–32.

[2] Achim Buerkle, Harveen Matharu, Ali Al-Yacoub, Niels Lohse, Thomas Bamber, and Pedro Ferreira. 2022. An adaptive human sensor framework for human–robot collaboration. *The International Journal of Advanced Manufacturing Technology* (2022), 1–16.

[3] Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems* 32 (2019).

[4] J. R. Comstock and R. J. Arnegard. 1992. *The Multi-Attribute Task Battery for operator workload and strategic behavior research.* Technical Report NASA Tech. Memorandum 104174. NASA Langley Research Center.

[5] Mustafa Demir, Nathan J McNeese, and Nancy J Cooke. 2020. Understanding human-robot teams in light of all-human teams: Aspects of team interaction and shared cognition. *International Journal of Human-Computer Studies* 140 (2020), 102436.

[6] Aaron Fisher, Cynthia Rudin, and Francesca Dominici. 2019. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of machine learning research: JMLR* 20 (2019).

[7] S. Fuchs and J. Schwarz. 2017. Towards a dynamic selection and configuration of adaptation strategies in Augmented Cognition. In *International Conference on Augmented Cognition*. Springer, 101–115.

[8] Ronnie J. Glavin. 2011. Human performance limitations (communication, stress, prospective memory and fatigue). *Best Practice & Research Clinical Anaesthesiology* 25, 2 (2011), 193–206. https://doi.org/10.1016/j.bpa.2011.01.004 Safety in Anaesthesia.

[9] J. Heard, J. Fortune, and Julie A Adams. 2020. SAHRTA: A Supervisory-Based Adaptive Human-Robot Teaming Architecture. In *IEEE Conference on Cognitive and Computational Aspects of Situation Management*.

[10] J. Heard, R. Heald, C. E. Harriott, and J. A. Adams. 2019. A Diagnostic Human Workload Assessment Algorithm for Supervisory and Collaborative Human-Robot Teams. *ACM Transactions on Human-Robotic Interaction* 8, 2 (2019), 1–30.

[11] Antonio R Hidalgo-Muñoz, Adolphe J Béquet, Mathis Astier-Juvenon, Guillaume Pépin, Alexandra Fort, Christophe Jallais, Hélène Tattegrain, and Catherine Gabaude. 2019. Respiration and heart rate modulation due to competing cognitive tasks while driving. *Frontiers in Human Neuroscience* 12 (2019), 525.

[12] Hye-Geum Kim, Eun-Jin Cheon, Dai-Seg Bai, Young Hwan Lee, and Bon-Hoon Koo. 2018. Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry investigation* 15, 3 (2018), 235.

[13] Laura Leuchs, Max Schneider, Michael Czisch, and Victor I Spoormaker. 2017. Neural correlates of pupil dilation during human fear learning. *Neuroimage* 147 (2017), 186–197.

[14] Marta Lorenzini, Wansoo Kim, Elena De Momi, and Arash Ajoudani. 2019. A new overloading fatigue model for ergonomic risk assessment with application to human-robot collaboration. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 1962–1968.

[15] Debasmita Mukherjee, Kashish Gupta, Li Hsin Chang, and Homayoun Najjaran. 2022. A Survey of Robot Learning Strategies for Human-Robot Collaboration in Industrial Settings. *Robotics and Computer-Integrated Manufacturing* 73 (2022), 102231.

[16] Likhitha Nagahanumaiah, Saurav Singh, and Jamison Heard. 2022. Diagnostic Human Fatigue Classification using Wearable Sensors for Intelligent Systems. In *2022 17th Annual System of Systems Engineering Conference (SOSE)*. IEEE, 424–429.

[17] Celal Savur, Shitij Kumar, and Ferat Sahin. 2019. A framework for monitoring human physiological response during human robot collaborative task. In *2019 IEEE international conference on systems, man and cybernetics (SMC)*. IEEE, 385–390.

[18] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (2021), 102551.

[19] Saurav Singh and Jamison Heard. 2022. A Human-Aware Decision Making System for Human-Robot Teams. In *2022 17th Annual System of Systems Engineering Conference (SOSE)*. IEEE, 268–273.

[20] Saurav Singh and Jamison Heard. 2022. Human-aware reinforcement learning for adaptive human robot teaming. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 1049–1052.

[21] Saurav Singh, Panos P Markopoulos, Eli Saber, Jesse D Lew, and Jamison Heard. 2023. Measuring Modality Utilization in Multi-Modal Neural Networks. In *2023 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 11–14.

[22] Feiyu Xu, Hans Uszkoreit, Yangzhou Du, Wei Fan, Dongyan Zhao, and Jun Zhu. 2019. Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8*. Springer, 563–574.