Interactively Explaining Robot Policies to Humans in Integrated Virtual and Physical Training Environments

Peizhu Qian
Rice University
Houston, TX, USA
pqian@rice.edu

Vaibhav Unhelkar
Rice University
Houston, TX, USA
vaibhav.unhelkar@rice.edu

ABSTRACT

Policy summarization is a computational paradigm for explaining the behavior and decision-making processes of autonomous robots to humans. It summarizes robot policies via exemplary demonstrations, aiming to improve human understanding of robotic behaviors. This understanding is crucial, especially since users often make critical decisions about robot deployment in the real world. Previous research in policy summarization has predominantly focused on simulated robots and environments, overlooking its application to physically embodied robots. Our work fills this gap by combining current policy summarization methods with a novel, interactive user interface that involves physical interaction with robots. We conduct human-subject experiments to assess our explanation system, focusing on the impact of different explanation modalities in policy summarization. Our findings underscore the unique advantages of combining virtual and physical training environments to effectively communicate robot behavior to human users.

CCS CONCEPTS

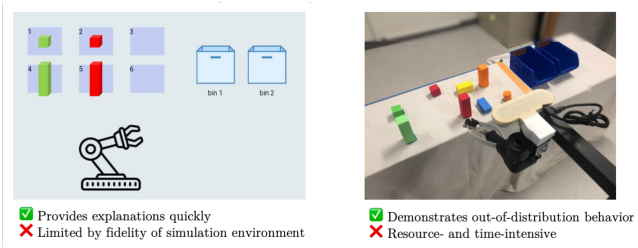
• **Human-centered computing** → **Interaction devices**; • **Computer systems organization** → **Robotics**.

KEYWORDS

Explainable AI; Value Alignment; AI-Assisted Human Training

ACM Reference Format:
Peizhu Qian and Vaibhav Unhelkar. 2024. Interactively Explaining Robot Policies to Humans in Integrated Virtual and Physical Training Environments. 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/10.1145/3610978.3640656

(a) A robot assisting a nurse. To effectively use this robot, the nurse needs an accurate mental model of robot behavior. We study training approaches that can help users acquire these mental models.



(b) Relative strengths of virtual and physical training.

Figure 1: We present an interactive policy summarization system that integrates virtual and physical training, enabling users to predict a robot's potential successes or failures. For a video demo, visit: http://tiny.cc/aiteacher-hri24.

1 INTRODUCTION

Robots are supporting humans in a variety of domains. For example, disaster response agencies are integrating robots to safeguard human firefighters [5, 13], and medical centers are experimenting with robots to alleviate nurse workload and enhance patient care [4, 15, 18]. As we envision a future where robots undertake increasingly significant and complex tasks alongside humans, a pivotal question arises: *What level of understanding about robots do we need to effectively and safely coexist with them?* This inquiry is crucial because robots in real-world may err, resting the responsibility for robot deployment with humans [12, 14, 19, 21, 24].

To formalize this question, let us imagine a nurse supported by a robotic assistant (Fig. 1a). When asked to help by the nurse, the robot can assist in gathering supplies, rearranging items in a patient room, and disposing of waste. To complete a given task, the robot iteratively senses the environment using its sensor observations (denoted as z), estimates the context ($s \approx \hat{s}$) using its observations $\hat{s} = \phi(z)$, and then takes action a according to its context-dependent behavioral policy: $a = \pi(\hat{s})$. However, as the robot's sensors, context estimation, or the policy may be imperfect, the robot might act incorrectly in some scenarios. Consequently, the nurse must be able to predict the robot's potential successes and failures to determine when to rely on its assistance.

HRI'24 Companion, March 11-15, 2024, Boulder, Colorado, USA

ACM ISBN 979-8-4007-0323-2/24/03
https://doi.org/10.1145/3610978.3640656

The paradigm of *policy summarization* aims to endow human users (such as nurses) with this understanding by generating informative summaries of robot behavior. Existing techniques generate these summaries either computationally, by selecting salient examples of robot behavior [1, 7, 9, 17], or interactively, by providing users with mechanism to ask questions [6, 10, 17]. More recently, hybrid techniques that combine the two approaches have been proven to effectively improve user understanding of AI systems, while also being subjectively preferred by humans [17]. Typically, these summaries are presented during a pre-task training session, aiding users in forming accurate mental models of the robots. Despite their foundational nature, however, most existing work on policy summarization has focused on simulated agents and environments, rather than physical robots [20, 23].

Consequently, **while significant attention has been paid to computing these summaries, there is a gap in understanding the most effective ways to communicate them to users.** Informed by research on human-robot communication [3, 6, 11, 22], we advocate that effectively *conveying* informative summaries is as critical as the task of *computing* them. To address this gap, this work makes two contributions. In Sec. 3, we integrate a state-of-the-art policy summarization algorithm [17] with an interactive user interface to summarize the behavior of physical robots. **The integrated system interactively provides policy summaries via two explanation modalities – virtual and physical – and is demonstrated on a mobile manipulation task.** This demonstration also highlights robotics-focused challenges in policy summarization that are not readily evident in simulation [25]. Physical interaction offers users a more comprehensive experience of robot contexts and behaviors, but training in physical environments generally demands more time and resources. Hence, in Sec. 4, we report on human subject experiments that assess the role of explanation modality on humans' understanding of robot behavior.

2 RESEARCH SCOPE

In line with the theme of "HRI in the real world," we focus on robots that are trained to solve *sequential tasks* in controlled environments (e.g., laboratories) but will be deployed in more general settings (e.g., open-world environments). Consider, for instance, the sorting task illustrated in Fig. 1b. Here, the Stretch RE-1 robot [8] aims to sort different types of blocks on a table into two bins. The task involves six pick locations and two drop bins, with objects varying in color and size. The reward function, used for training, encodes the following preferences for pick-up: tall red > tall green > small red > small green, with the pick location used to break ties when multiple blocks of the same type are present on the table; and drop-off: tall blocks in bin 1 and small blocks in bin 2. At each step, the robot has to select from 9 actions: 6 pick actions corresponding to each pick location, 2 drop actions, and wait.

Further, we consider robots that act autonomously by first estimating the task state from its observation; and then selecting its actions based on the inferred task state. For example, in the sorting task, the task state (s) is defined as the object type at each pick location and that in the robot's gripper, resulting in $\approx 80k$ nominal states. Given the state space, action space, and reward function, we model the task as a Markov decision process (MDP) and use the value iteration algorithm to derive the robot policy $\pi(s)$ [16]. To execute this policy, the robot needs a state estimate $s \approx \hat{s}$. The robot uses its two depth cameras and proprioception to sense its environment z and estimate the state, $\hat{s} = \phi(z)$. In our implementation, this state estimation is done using a rule-based computer vision module. Recall that, during its real-world operations, the robot may encounter unexpected objects, which are not captured in its state representation. Together, ϕ and π generate the robot's behavior $a = \pi(\phi(z))$ in both nominal and unexpected scenarios, which we refer to as in-distribution (ID) and out-of-distribution (OOD) scenarios, respectively. Given the tasks, scenarios, and robot behavior, we consider the problem of **creating a training system that improves a user's ability to predict the robot behavior.**

3 POLICY SUMMARIZATION SYSTEM

To address this problem, we design, demonstrate, and evaluate an interactive policy summarization system. The system utilizes an existing algorithm to generate policy summaries. These summaries are then conveyed to humans via a novel, interactive user interface, which leverages both virtual and physical training environments.

3.1 Generating Policy Summaries

Policy summarization methods select example demonstrations of robot behavior, with the goal of enabling users to accurately predict the robot's actions during deployment. In our sorting task, this involves helping users understand the robot's decision-making process, such as which object it will pick next, by showing (s, a)-demonstrations of robot behavior. In practice, it is infeasible to demonstrate robot behavior in every possible state and, thus, algorithms are required to select a small number of informative examples. To generate informative examples, we apply a recent policy summarization technique called AI Teacher [17]. AI Teacher provides two types of examples: algorithmically-generated *Teacher's* examples and user-generated *Custom* examples.

To generate teacher's examples, AI Teacher models the user as a Bayesian learner, inspired by models of human cognition [2]. In particular, it assumes that the human maintains a set of hypotheses $e \in E$ regarding robot behavior. The explanation algorithm then selects (state, action)-tuples as demonstrations that most effectively reinforce the user's belief in the hypothesis e^*, corresponding to the robot's actual policy π. These are called the teacher's examples. For comprehensive details on this algorithm, please refer to [17]. Utilizing this algorithm, we generate teacher's examples for the sorting task. Our implementation defines E through 32 hypotheses, created by varying the prioritization of different item types and drop-off locations in the reward function. Each teacher's example is generated as a (s, a)-trajectory of length 6 and is followed by a question (called quiz) regarding robot behavior.

To generate custom examples, AI Teacher involves a virtual training environment. Within this environment, users can craft their own *custom* ID scenarios and subsequently request demonstrations of the robot's behavior in these settings. Although AI Teacher is originally designed for solely explaining ID behavior via virtual training, as explained next, we extend its application to OOD scenarios by considering physical training environments that allow for collocated human-robot interaction.

3.2 Communicating Policy Summaries

To explain the behavior of physical robots, we design an interactive user interface that seeks to combine the relative strengths of virtual and physical training. In particular, the interactive interface enables the users to select the explanation modality (virtual or physical), select teacher's examples, and design custom examples. Examples requested using the virtual modality are shown as animation, while those using the physical are shown on the Stretch RE-1 robot.

While AI Teacher is originally designed for explaining ID behavior, in our application, we use its custom examples also for explaining OOD behavior. This extension is made possible by leveraging the physical robot. By utilizing the physical environment while requesting custom examples, the user is not limited by the scope of a simulation and can truly create any OOD scenario of interest to learn about the robot's representation $\hat{s} = \phi(z)$ and behavior $a = \pi(\hat{s})$. As we find in our human subject experiments, users use this mechanism to create unexpected scenarios, which are difficult to capture in virtual training. Together the algorithm and the user interface complete the design of the XAI system to summarize policies of a physical robot.

4 METHODOLOGY

We now validate our policy summarization system and assess the role of explanation modality via two sets of human studies, approved by Rice University's IRB. To consider both robot- and human-centric elements, we formulate the following hypotheses:

- H1 *Irrespective of the explanation modality, users can predict in-distribution robot behavior with high accuracy after receiving explanations using policy summarization techniques.*
- H2 *Users that receive policy summaries via a physical robot outperform those that do not (control group) in predicting out-of-distribution robot behavior.*
- H3 *Users subjectively assess receiving policy summaries via a physical robot to be important for improving robot transparency.*

4.1 Pilot Study

First, we conduct a pilot study with the goal of validating the designed system and finalizing the design of a larger experiment. In this open-ended study, participants were provided the policy summarization system and asked to use it to understand robot behavior in the sorting task. Upon completing the training, they answer three sets of questions:

- Given an in-distribution scenario s, predict the robot action $a = \pi(s)$. We call these as *forward ID* questions.
- Given an action a and a set of conditions C, design a scenario that leads the robot to select the action under given conditions: $s \in S, s.t.(\pi(s) = a) \wedge (s \models C)$). For example, "Use at least 4 objects to create a scenario where the robot will pick up a tall red block from location 6." We refer to these as *inverse ID* questions.
- Given an out-of-distribution scenario with observation z, predict the robot's action $a = \pi(\phi(z))$. We refer to these as *forward OOD* questions.

We administer 35 questions, among which 20 are forward ID questions (worth 1 point each), 5 are inverse ID questions (2 points), and 10 are OOD questions (2 points). Some questions are given more points because they are perceived as harder questions. Altogether they add up to 50 points. We conduct this pilot study with 6 participants recruited from Rice University. We observe that participants score high ($M = 45.33/50, SD = 3.14$), thereby providing preliminary validation for the efficacy of the integrated XAI system across both ID and OOD behavior. Participants point out that the virtual robot is more efficient for understanding in-distribution behavior and, thus, helps them quickly build a rough understanding of the robot. On the other hand, participants understand that while the physical robot takes longer to complete the same actions, it can help confirm the consistency of simulation, test the robot's sensors, and learn about out-of-distribution behavior.

4.2 Experiment Design

Next, to assess the relative strengths of virtual and physical training environments, we conduct a between-subject randomized controlled trial with one independent variable: explanation modality. The robot, experimental task, explanation algorithm, and user interface remain identical to the pilot study. The control group uses only the virtual environment as the explanation modality. The experimental group, similar to Sec. 4.1, is given access to both the physical and virtual environments to receive summaries.[1]

4.2.1 Procedure. The experiment takes place in a laboratory. The session starts with a briefing from the supervisor on the purpose, procedure, and participant's rights. After giving written consent, participants complete a demographic survey. Next, the participants complete the supervisor-guided training task to familiarize themselves with the integrated XAI system. At the end of the session, which lasts around 40 min, the participants are thanked for their participation and receive a gift card of $10.

4.2.2 Dependent Measures. Similar to Sec. 4.1, we utilize an objective test to assess participants' understanding of robot behavior. The test included 10 Forward ID questions, 5 Inverse ID questions, and 15 Forward OOD questions. To better tease out the difference in user understanding across ID and OOD behavior, we increase the proportion of OOD questions from the pilot study but give all questions even points (1 point per question) regardless of question type to reduce confounding effects. To design OOD questions that reflect real-world situations, we consider the following cases: placing multiple objects at one pick location (the robot assumes at most one object per location), placing objects in unexpected locations, using unexpected objects, and changing the room lighting. All ID questions are administered using the virtual modality, while OOD questions using the physical robot. Further, the participants complete a post-experiment survey, which asks them about their learning experience as well as the role of the virtual and physical training environments.

4.3 Results and Discussions

We recruit 24 participants (13 female, 11 male) from Rice University. Participants' age ranged between 21 and 39 years.

[1]We considered an alternative treatment, which *only* uses the physical modality. However, based on the pilot study, we ascertained that the virtual modality is critical for learning ID behavior and hence make it available to both the groups.

Table 1: Participants' performance on the test assessing their prediction of robot behavior in the sorting task.

	ID (%)	OOD (%)	All Questions (%)
Control	97.22	75.00	86.11
Experimental	100.00	**81.11**	90.56

Table 2: Average learning time and instructions used by participants to learn robot behavior in the sorting task.

	Time (min)		Examples (#)	
	Virtual	Physical	Virtual	Physical
Control	3.4	-	46.5	-
Experimental	5.3	**8.2**	53.7	**10.2**

Result 1. Participants demonstrate an incredible ability to understand in-distribution robot behavior from a small number of examples. Participants in both groups score over 97% (Table 1, *ID*) in predicting the robot's ID behavior from only seeing less than 0.1% (Table 2, *Examples*) of the ID states. We run an one-sided Wilcoxon signed-rank test and the result show that the median score on ID questions is higher than 80%[2], verifying H1 ($p < 0.001$).

Result 2. Using the physical modality, participants can marginally better predict the out-of-distribution robot behavior. Unsurprisingly, participants find it more challenging to predict out-of-distribution robot behavior relative to in-distribution behavior. Nonetheless, the participants in both groups score on average 75% (Table 1, *OOD*) or higher on the OOD questions. To test Hypothesis H2, we conduct a Wilcoxon rank-sum test to evaluate the effect of different treatments. While we observe an improvement using the physical modality, the effect size is marginal and not statistically significant ($p = 0.20$). One possible explanation for this observation is that the participants take accurate, informed guesses of how the robot will treat unseen objects (i.e., the robot will ignore these objects).

Result 3. Participants perceive the physical and virtual modality as equally important and prefer an integrated XAI system that includes both modalities. Although we do not see a significant difference in participants' objective performance, we observe that participants subjectively perceive both modalities as of high and equal importance. In the post-experiment survey, we ask participants to rate the importance of the virtual robot and physical robot for their learning of the robot behavior, respectively. Of the 24 participants, 10 participants rate the physical robot to be extremely important, 9 participants as very important, and 5 participants as important. These scores give the physical robot an average importance of 6.21 (out of 7), thereby providing evidence in support of Hypothesis H3. As one of the participants writes in the survey,

> *The ability to try many different scenarios quickly seems to be the most satisfying thing when learning about the robot behavior. Since the physical robot is noticeably slower than the virtual robot, I feel that the virtual robot is much more satisfying to a user who wants to learn about the robot behavior quickly. On the other hand, the physical robot displays failure modes that the virtual robot does not, so perhaps the user's perception of the virtual robot as "better" is incorrect in real-world usage.*

In a follow-up question, we ask participants to indicate their preferred allocation of training time with each robot type. No participant selects to only use one modality; instead, participants prefer different ratios of training with each modality.

The types of custom scenarios that participants design using the physical robot are also informative. For instance, we observe one participant use different green-colored objects available to them (a green pack of gums and green markers instead of green blocks) to assess whether the robot will pick them. Another participant considers the case where an unexpected object (in their case, the participant uses their shoe!) is placed on the table and requested indirect instructions in this unusual scenario. These unusual stimuli presented to the physical robot by participants suggest that they understand that a robot can encounter OOD scenarios during real-life deployment, which cannot be tested in the virtual simulation.

Result 4. Participants request a majority of examples using the virtual robot but spend the majority of learning time with the physical robot. Table 2 summarizes the average time spent by participants with the summarization system and the number of exemplary demonstrations requested. Across both the control and experimental groups, the participants request ≈ 50 examples using the virtual training environment. Participants in the experimental group, additionally request 10 demonstrations using the physical robot.

5 CONCLUSION

Our work is an initial investigation in applying policy summarization techniques to explaining behavior of physical robots. Towards this effort, we develop an interactive policy summarization system that utilizes virtual and physical training environments. We demonstrate the system on a robotic sorting task (with $\approx 80k$ states and a complex reward structure) and evaluate it via human subject experiments. Our experiment results, which demonstrate the utility of policy summarization and the relative strengths of the two training environments, are relevant for explainable AI and robotics research as well as for practitioners who train humans to use robots.

Our work also offers several avenues for future work. While we take steps to ensure that the experimental task is sufficiently complex, we encourage reproductions of our work using other robotics tasks. Second, to avoid experimental confounds, we strive for consistency in robot behavior across the virtual and physical training environments. However, in real-world HRI, the sim-to-real gap (differences due to sensor noise or actuator variability) across training environments may impact human understanding of robot behavior. We suggest future research to examine how such discrepancies impact users' ability to understand robot behavior. Lastly, our work highlights the need for algorithmic techniques that can explain both in- and out-of-distribution robot behaviors.

ACKNOWLEDGMENTS

We thank Bryant Cassady for assistance with the robotics implementation. Peizhu Qian was supported in part by the ARO CA# W911NF-20-2-0214, NSF award# 2222876, and Rice University funds.

[2]The 80% threshold is informed by performance in similar simulated tasks [17].

REFERENCES

[1] Dan Amir and Ofra Amir. 2018. HIGHLIGHTS: Summarizing Agent Behavior to People. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[2] Chris Baker, Rebecca Saxe, and Joshua Tenenbaum. 2011. Bayesian Theory of Mind: Modeling Joint Belief-Desire Attribution. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

[3] Tathagata Chakraborti, Sarath Sreedharan, Yu Zhang, and Subbarao Kambhampati. 2017. Plan Explanations as Model Reconciliation: Moving beyond Explanation as Soliloquy. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

[4] Eftychios G. Christoforou, Sotiris Avgousti, Nacim Ramdani, Cyril Novales, and Andreas S. Panayides. 2020. The Upcoming Role for Nursing and Assistive Robotics: Opportunities and Challenges Ahead. *Frontiers in Digital Health* (2020).

[5] The Los Angeles City Fire Department. 2020. LAFD debuts the RS3: First Robotic Firefighting Vehicle in the United States. https://www.lafd.org/news/lafd-debuts-rs3-first-robotic-firefighting-vehicle-united-states. Accessed: 2023-11-30.

[6] Bradley Hayes and Julie A. Shah. 2017. Improving Robot Controller Transparency Through Autonomous Policy Explanation. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[7] Sandy H. Huang, Kush Bhatia, P. Abbeel, and Anca D. Dragan. 2018. Establishing Appropriate Trust via Critical States. *IROS* (2018).

[8] Charles C Kemp, Aaron Edsinger, Henry M Clever, and Blaine Matulevich. 2022. The Design of Stretch: a Compact, Lightweight Mobile Manipulator for Indoor Human Environments. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*.

[9] Michael S. Lee, Henny Admoni, and Reid Simmons. 2021. Machine Teaching for Human Inverse Reinforcement Learning. *Frontiers in Robotics and AI* (2021).

[10] Meghann Lomas, Robert Chevalier, Ernest Cross II, Robert Garrett, John Hoare, and Michael Kopack. 2012. Explaining Robot Sctions. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[11] Nikolaos Mavridis. 2015. A Review of Verbal and Non-Verbal Human-Robot Interactive Communication. *Robotics and Autonomous Systems* (2015).

[12] Donald A. Norman. 1990. The 'Problem' with Automation: Inappropriate Feedback and Interaction, Not 'Over-Automation'. *Philosophical transactions of the Royal Society of London.* (1990).

[13] Liubove Orlov-Savko*, Zhiqin Qian*, Gregory M Gremillion, Catherine E Neubauer, Jonroy Canady, and Vaibhav Unhelkar. 2024. RW4T Dataset: Data of Human-Robot Behavior and Cognitive States in Simulated Disaster Response Tasks. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[14] Raja Parasuraman and Victor Riley. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors* (1997).

[15] Joseph Andrew Pepito, Hirokazu Ito, Feni Betriana, Tetsuya Tanioka, and Rozzano C Locsin. 2020. Intelligent Humanoid Robots Expressing Artificial Human-like Empathy in Nursing Situations. *Nursing Philosophy* 21, 4 (2020).

[16] Martin L Puterman. 2014. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.

[17] Peizhu Qian and Vaibhav Unhelkar. 2022. Evaluating the Role of Interactivity on Improving Transparency in Autonomous Agents. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*.

[18] Carlos Quintero-Pena*, Peizhu Qian*, Nicole M Fontenot, Hsin-Mei Chen, Shannan K Hamlin, Lydia E Kavraki, and Vaibhav Unhelkar. 2023. Robotic Tutors for Nurse Training: Opportunities for HRI Researchers. In *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*.

[19] Yao Rong, Tobias Leemann, Thai-Trang Nguyen, Lisa Fiedler, Peizhu Qian, Vaibhav Unhelkar, Tina Seidel, Gjergji Kasneci, and Enkelejda Kasneci. 2023. Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).

[20] Tatsuya Sakai and Takayuki Nagai. 2022. Explainable Autonomous Robots: a Survey and Perspective. *Advanced Robotics* (2022).

[21] Thomas B. Sheridan and Raja Parasuraman. 2005. Human-Automation Interaction: Taxonomies and Qualitative Models. *Reviews of Human Factors and Ergonomics* (2005).

[22] Vaibhav V Unhelkar, Shen Li, and Julie A Shah. 2020. Decision-making for Bidirectional Communication in Sequential Human-Robot Collaborative Tasks. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[23] Sebastian Wallkötter, Silvia Tulli, Ginevra Castellano, Ana Paiva, and Mohamed Chetouani. 2021. Explainable Embodied Agents Through Social Cues: a Review. *ACM Transactions on Human-Robot Interaction (THRI)* (2021).

[24] X Jessie Yang, Vaibhav V Unhelkar, Kevin Li, and Julie A Shah. 2017. Evaluating Effects of User Experience and System Transparency on Trust in Automation. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*.

[25] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. 2020. Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey. In *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)*.