Development and Evaluation of a Deep Q-Network-Based Robot Learning Paradigm in Real-World Human-Robot Collaborative Tasks

Garrett Modery, Weitian Wang*, *Senior Member, IEEE*, Rui Li, Yi Chen, and Mengchu Zhou, *Fellow, IEEE*

***Abstract* — As robot systems continue to be advanced and implemented across industries, they do so typically by two methodologies, including standalone systems and collaborative ones. Standalone systems are typically set in their own areas, away from human workers. Collaborative robots share a common workspace with their human counterparts and work with them to complete tasks together efficiently and safely. Within this category of robotics, there exists another subcategory that describes the method of implementation and usage rather than simply the type of system. This subcategory involves how the machine will interact with workers and understand its role in interaction. Thus, it raises interest in the field of learning from demonstrations, where the robot may dynamically learn the behavior that is desired by the user rather than being explicitly hardcoded to perform its task. In this work, we develop a deep Q-network-based robot learning paradigm for human-robot partnerships in shared tasks. The proposed approach is validated in real-world human-robot collaborative contexts. In addition, to assess the performance of this approach and the acceptance from practitioners, we conduct a multi-metric user study. Implementation and evaluation results indicate that the developed solution works effectively for human-robot teamwork and receives high support from active users who rate it very well on several key metrics. The future work of this study is also discussed.**

I. Introduction

As Industry 4.0 continues to revolutionize manufacturing sectors, new technologies in line with its values are consistently being produced. These values include further automation of factories, the implementation of smart systems within them through models such as Industrial IoT, and an overall transfer of some autonomy to machines, facilitated by information systems [1-3]. Industry 4.0 was the natural step in the sector, making use of new, versatile technologies to enhance productivity. What it does *not* place emphasis on, however, is the human worker who will inevitably be working within these factories. Now, with the approaching rise of Industry 5.0, human-centricity becomes a focal point [4].

Though research into Industry 5.0 is admittedly scarce, the three pillars that uphold it are well-defined: sustainability, resiliency, and human-centricity. Sustainability refers to not only the reduction in environmental impact, but also the efficient use of natural resources and support of an effective economy. Resiliency involves the robustness of systems in the face of disruptions and the ability to recover quickly to a stable state during and after geopolitical shifts. Finally, the most pertinent for this research is the focus of human-centricity. Human-centricity could be defined as an approach that places human interests and needs at the center of the production process [5]. Lu *et al.* noted that future human-machine teams will need to place these values at the center of manufacturing planning and control [6]. This places much emphasis on adaptive robotic systems that work with people fluently. Systems that are capable of adaptation to their work environment, as well as their companions, are crucial for supporting the human-centric vision of Industry 5.0. In human-robot collaborative contexts, the collaborative machines will be used for their strengths, such as performing repetitive and labor-intensive tasks, while humans will exercise their strengths, such as critical thinking and personalization [7, 8]. These collaborative machines are specifically designed to accommodate human partnerships, reinforcing their role in the human-centric focus of Industry 5.0.

It is the purpose of this work to develop a robot learning paradigm and evaluate the acceptance of one such adaptive robotic system by human participants in human-robot collaborative contexts through a user study. As humans enter a place of greater importance in the human-robot dynamic, it is only natural that systems must exist to accommodate them to the best of their ability, and these humans accept the system itself. These systems should be adaptable, but also offer fluent and flexible interaction so as not to harm the user's satisfaction nor their productivity. We present a reinforcement learning-based human-robot collaborative approach and a user study to evaluate it. This work seeks to identify if this is a solution to human-centered human-robot collaboration (HRC) that has potential in the industry through the collection of subjective ratings of participants. These participants interact with the robot in a variable-length collaborative assembly task and, through a multi-metric survey, provide feedback on the system as well as the interaction as a whole by several key evaluation indicators. This work seeks to determine from the data how acceptable users of the system find it and potentially establish areas of necessary improvement.

II. Related Work

Reinforcement learning (RL) has been used in several robotics applications, such as object grasping [9], in which the robots consistently attempt to perform a grasping action and improve their policy from that data. This is primarily robot-oriented, meaning that after taking an action, the robot examines the quality of that action and handles its policy adjustments without outside intervention. It has been applied to collaborative robotics as well, where Gomes *et al.* presented that the working environment of such robots is often subjected to unforeseen modifications by people [10]. While reinforcement learning may be utilized for enhancing the grasping capabilities of robots, that is not the only extent of its applications. Multi-agent systems, such as in [11], are more in line with a collaborative system. In this work, the collaborative task involved a ball in a maze game that required a human to control one axis of rotation and a robot to control the other, demanding cooperation and giving the agent the opportunity to learn from its partner. Notably, the authors

G. Modery, W. Wang, and R. Li are with the School of Computing, Montclair State University, Montclair, NJ 07043 USA. (corresponding author: wangw@montclair.edu)

Y. Chen is with ABB Corporate Research Center, Raleigh, NC 27606 USA.

M. Zhou is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA.

noted that there are certain complications that come from bringing humans into the operation (e.g., unpredictability). This issue highlights the importance of intention prediction, which is a component that is considered in this study and will be discussed in greater detail in the following sections.

Reinforcement learning demands a significant number of experiences to learn from teachers to reach the optimal policy [12]. As it takes humans practice to reach high performance in a task, reinforcement learning agents are no different. In virtual environments where agents are taught to play games, one can simply let the simulation continue without many delays. In robotics, however, there exist several issues as described in detail in [13], though perhaps most obvious is the delays in sensors and actuators of the robot that is training. For this reason, it is common to use simulated environments to cut down on the time required for robot training in the real world, even if they do not perfectly represent the real-world conditions that the agent will ultimately operate in [10]. To mitigate these gaps, we make use of the real-world human-robot collaborative environment in this study to gather subjective ratings of such an RL collaborative system.

III. TLPC Framework

This paradigm builds on our prior studies in the development and refinement of a Teaching-Learning-Prediction-Collaboration (TLPC) framework, which is designed to permit robots to learn from and collaborate with humans [14, 15]. This dynamic system is created to accommodate personalized tasks that may vary between users. As research continues, the particular operations in each phase of this framework may shift, though the overarching theme is the same.

In this framework, the components of human-robot collaboration are broken down into distinct phases, including teaching, learning, prediction, and collaboration. The teaching phase involves a human operator demonstrating the task being completed in a particular way to the robot. The specific manner of task completion may differ from user to user, and as such furthers the importance of permitting customizability in the interaction. Following the teaching stage, the learning phase involves the robot using its collected task knowledge and applying it through a generalized method, such as state machines. In the prediction and collaboration phases, the robot examines its environment's current states to predict the human's next action with its learned knowledge. This prediction permits the robot to take action to collaborate and assist the user in the completion of their task. For the TLPC framework, specific implementations vary in their execution of each phase. In this study, we will develop a deep Q-network-based robot learning approach within TLPC for human-robot teams in real-world collaborative tasks and evaluate the proposed solution via a user study.

IV. Approaches

A. Deep Q-Learning

Reinforcement learning is a paradigm designed to enable active learning by an agent via permitting it to take actions within its environment and examine the quality of those actions to learn and optimize its policy. The selection of these actions raises the prediction problem [16]. That is, to determine the state or action-value function for a policy. The primary focus of reinforcement learning, from a high-level view, is to enable the agent to learn the optimal policy for sequential decision problems and, in doing so, maximize its rewards for a given task. In essence, RL involves an agent – the executor of selected actions – operating within an environment. This environment exists in a current state at a given time t, which the agent observes and considers to select its next action. This state, s_t from state space S is input into the prediction model and, by following policy $\pi(a_t \mid s_t)$, produces an action a_t from action space A [16]. Upon taking the action a_t, the environment transitions into a state s_{t+1} and the environment returns a scalar reward r_t given by the reward function $R(s, a)$. Additionally, in an episodic setup, a terminal state d is eventually reached, at which point it restarts. The experiences from which the reinforcement occurs are stored as a 5-tuple:

$$(s_t, a_t, r_t, s_{t+1}, d)\ . \tag{1}$$

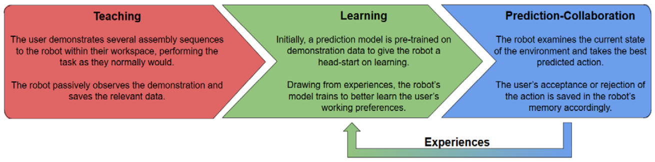
Deep Q-learning is an extension of traditional reinforcement learning methods that utilizes a deep neural network to address the prediction problem by approximating the Q-values for a given state. These Q-values are approximated using the Bellman equation [17]:

$$Q(s, a) = r(s, a) + \gamma \max Q(s', a')\ . \tag{2}$$

In the Bellman equation, the Q-value for state action pair (s, a) is calculated as the sum of the immediate reward gained by taking action a while in state s and the maximum reward for the next state s', over all possible actions a'. The discount factor γ is applied to the latter half of the equation to offset the value of future rewards compared to immediate ones, defined as $\gamma \in (0, 1]$. In our implementation, we utilize a double Q-learning setup [18]. That is, using one main network to predict the Q-values of a given state and another target network for stability by using it to calculate the loss at each step.

B. TLPC Framework Deep Q-Network Implementation

The particular details of the implementation of the TLPC framework will vary depending on the approach, and this is especially true of the learning phase. Variability in human execution of tasks is difficult for deterministic models to adapt to, and as such, demands a dynamic system that can circumvent this barrier. Thus, in order to maximize the effectiveness of the human-robot partnership, an active learning agent, such as one driven by reinforcement learning, is a promising option [6]. Additionally, Akalin *et al*. remarked that many human-robot interactions may be structured as sequential decision-making tasks which, by nature, are typical RL problems [19].



The user demonstrates several assembly sequences to the robot within their workspace, performing the task as they normally would.

The robot passively observes the demonstration and saves the relevant data.

Initially, a prediction model is pre-trained on demonstration data to give the robot a head-start on learning.

Drawing from experiences, the robot's model trains to better learn the user's working preferences.

The robot examines the current state of the environment and takes the best predicted action.

The user's acceptance or rejection of the action is saved in the robot's memory accordingly.

Fig. 1. The TLPC framework deep Q-network implementation.

This work adapts the TLPC framework to accommodate active learning by enabling the collaborative robot to receive and apply real-time feedback from its human partner through a deep Q-learning algorithm. Fig. 1 outlines the TLPC framework deep Q-network (DQN) implementation. First, the teaching stage involves the user demonstrating assembly sequences to the observing robot within their workspace. These sequences are of variable length and are structured according to their preferences. The demonstrated sequences are denoted as ζ, where individual sequences are denoted as ζ_i.

The learning phase of this framework breaks the sequence into states, or experiences, which the prediction model may train with. States within this model, *s*, are represented as matrices of shape (*LT, NA*), where *LT* is the maximum length of a given task (though there is theoretically no limit), and *NA* is the number of possible actions represented as one-hot encoded vectors that are the length of the action space and are by default initialized to zero. The states of a given human-robot collaborative task can be expressed as follows:

$$s = \left[s_1, s_2, \cdots, s_{LT}\right]^{\mathrm{T}}, \tag{3}$$

$$s_i = [0, 0, \cdots, 1, \cdots, 0] \text{ for } i = 1, \cdots, LT. \tag{4}$$

Next, the demonstration data is acquired as transitions within the replay buffer β. For each sequence μ at ζ_i, transitions τ are created according to the 5-tuple structure used in reinforcement learning setups, which can be generated by:

$$\tau_j = \{\sigma(\mu_{0:j}), v(\mu_{j+1}), r, \sigma(\mu_{0:j+1}), d\}. \tag{5}$$

Note that σ represents the function that turns the sequence of observed components into the state representation described above, *v* is the one-hot vector encoding function, *r* is the reward given by the reward function, and *d* is the flag that indicates whether this transition leads into a done state.

With the creation of transition τ_j, it can be added to the replay buffer:

$$\beta = \beta U\{\tau_j\}. \tag{6}$$

The prediction model is pre-trained with these transitions created from the demonstration data to permit faster convergence to the optimal policy and to reduce the amount of direct feedback required by the user. This pre-training makes use of four loss types, including the one and n-step double Q-learning losses, a large margin classification loss, and an L2 regularization loss. The one and n-step double Q-learning losses are forms of temporal difference (TD) loss, and they use the Huber loss function in their implementation. This function utilizes both mean squared error and mean absolute error [17], which is defined as:

$$L(\delta, y, f(x)) = \begin{cases} \frac{1}{2}(f(x)-y)^2 & \text{if } |f(x)-y| \le \delta, \\ \delta|f(x)-y| - \frac{1}{2}\delta^2 & \text{if } |f(x)-y| > \delta. \end{cases} \tag{7}$$

The Huber loss function, L, is used for its suitability in handling outliers in data with these two error types. It does so by making use of the δ parameter, which establishes the threshold for switching between the two components of the function. Additionally, *y* is the target value, and *f(x)* is the predicted value. The large-margin classification loss forces the values of non-demonstrator actions to be at least a margin lower than the value of the demonstrator's action, effectively prioritizing them above all others [20]. which is defined as:

$$J(Q) = \max[Q(s, a) + l(a_E, a)] - Q(s, a_E), \tag{8}$$

where a_E represents the action of the demonstrator, and $l(a_E, a)$ is the margin function that is positive if $a \neq a_E$, and 0 if $a = a_E$. Finally, the L2 regularization loss is applied to prevent overfitting on the demonstration data.

Simulated environments are used for the application of policy optimization, in which the agent may explore actions within its action space and thus build its experiences from which it trains. Positive rewards are given for actions that align with the demonstration data, and negative ones are given for those that do not. These simulated experiences may not entirely represent the desired behavior of the robot, and as such, experiences that the robot takes within the physical environment have their rewards weighted differently. Following this process and with learning completed, the robot examines the state of the assembly process, then makes a prediction of the next state and takes action to retrieve the appropriate next part for the operator. Thus, the robot enters the prediction-collaboration phase. Should the robot make the incorrect selection for its human partner, the user will reject the part and place it back into the secondary workspace. The rejection of the selection will indicate to the robot that this state transition was improper, and the experience will be collected into the robot's memory as such with a negative reward associated with it. Alternatively, upon making the correct selection, the user will accept the part into their primary workspace and the state transition will be considered proper and will be held appropriately with a positive reward. This process is presented in Fig. 2 below.

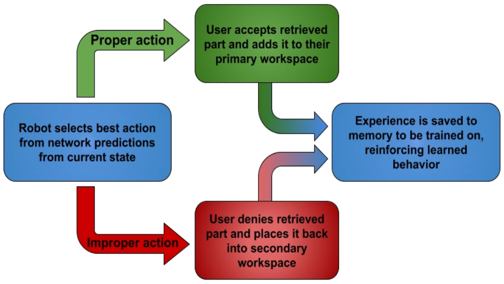

Proper action
User accepts retrieved part and adds it to their primary workspace
Robot selects best action from network predictions from current state
Experience is saved to memory to be trained on, reinforcing learned behavior
Improper action
User denies retrieved part and places it back into secondary workspace

Fig. 2. Human-robot feedback process.

Naturally, all of these experiences, both proper and improper, will be used to enhance the robot's operation. Sampling the replay buffer for acquired experiences, the robot briefly regresses into the learning phase of the TLPC framework to reinforce its prediction model before returning to the prediction-collaboration stage to continue working with the user. In this way, an active learning structure is modeled that enables real-time feedback and learning during the interaction. Additionally, though the robot begins with the demonstration data, it serves only as a starting point for the model. That being the case, a human operator need not be a professional at their task, as they may reject and approve selections as they see fit, and the robot will continue to advance in its abilities and knowledge through this feedback.

V. User Study Design

A. Experimental Platform

A user study is conducted to evaluate the performance of the proposed approach and the acceptance of a human-robot collaborative system that learns and operates through the reinforcement learning paradigm. It makes use of the following equipment: a Franka robot, an Intel RealSense D435i camera mounted to the end-effector of the robot, two Lenovo P520 ThinkStations, and fifteen wooden parts with letters on. The parts are used to represent assembly tasks for participants engaging in the study. One ThinkStation runs the MoveIt Commander Python API to plan and execute the robot's movement trajectory commands as well as to run the developed robot learning algorithm, and the other hosts the survey that is connected to the database where the results are anonymously collected. The RealSense camera provides the robot with a view of its workspace and the participant's selection area. Fig. 3 shows the experimental setup.

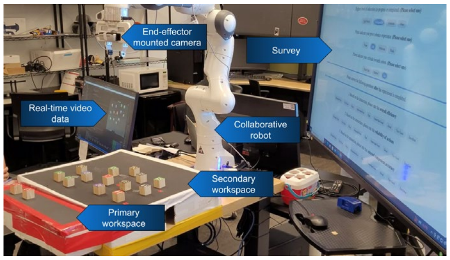


Fig. 3. Experimental setup [21].

B. Task Design

The task design for the user study is as follows. Participants of the study stand across from the robot with the end effector in the "hand-off" state, which permits the camera to view the primary workspace, or selection area, of the user. One sequence at a time, they position the parts in a particular order of any length that they desire, producing three orderings for the robot to learn. Participants are encouraged to keep the sequences similar enough so as to demonstrate the approaches' capability to adapt to such scenarios, though such a choice is not necessary and entirely different sequences can be used. For example, consider the selection of two orderings of the parts: ROBOT and ROMOT. When the robot examines the state RO, the deep Q-learning approach to the issue encourages the robot to make the decision between which action, B or M, will ultimately produce the maximum reward. The more often a sequence is employed during interaction, the more this knowledge will be reinforced, and thus the robot will be more confident in its prediction of these transitions.

After the sequences are demonstrated to the robot, it will enter its learning phase in which it pretrains briefly on the state transitions produced from the demonstration data following the learning process outlined in section IV.B. During this time, participants are instructed to take the first one or few parts of their sequence and add them to their primary workspace. Once the robot completes its learning process, it examines the starting state and begins its prediction-collaboration phase from there. To extend the previous example, consider that, while the robot is learning, the participant selects R from the secondary workspace and adds it to theirs. Once it finishes, the robot will select and execute the best-predicted action – in this case, O – and retrieve that component for the user. Alternatively, starting with RO is also a possibility, and the robot will pick up wherever the task begins. This ability to "jump into" any state that the environment may be in is one of the many valuable qualities of the reinforcement learning approach, especially for real-world contexts.

C. Data Collection

The data collected from this user study is gathered from a survey provided at the start and end of the experiment. We recruit 21 participants for this study (8 are female and 13 are male). Prior to interacting with the robot, participants are asked to answer two questions, including prior robotics experience and attitude towards robots. We request a self-evaluation from participants regarding their prior experience with robots, scaled from "none" to "much" on a four-point scale, as well as their attitude towards robots on a scale from "don't like" to "like" on a three-point scale. Both metrics are selected to gather background information from the participant that may help to gather insights into potential biases of their ratings of the interaction, particularly the rating of trust [22].

The post-interaction survey questions that are asked of participants include overall efficiency, reliability of actions, selected component accuracy, learning speed, adaptability to task, perceived safety, overall comfort, and trust. These questions are offered on a nine-point Likert scale [15]. We allow participants to evaluate the system's efficiency after the collaboration. This is done first with a direct question, but also with more specific queries to help identify the reasoning behind the selection. These queries are the reliability of actions metric, the selected component accuracy metric, and the learning speed metric. Participants' ratings of the robot's adaptability to the task are also collected. Safety and comfort are also evaluated, and these are often related to the metric of trust [23]. In the context of this study, trust defines how likely an agent (the participant) is to trust another (the robot) to perform its duty based on its previous activity. By evaluating the metrics of safety and comfort, trust may be better understood.

D. Methods of Data Analysis

We evaluate the survey results in two ways. First, we compare three metrics (comfort, safety, and efficiency) that were used in our previous work, which utilized a state machine approach for modeling learning and collaboration in a similar assembly task [15]. From this, we seek to determine the difference in responses, if any, between the two approaches. In order to perform these evaluations, we make use of the two-tailed Brunner-Munzel (BM) test, which is a nonparametric rank-based test that is considered to be more robust when compared to other rank-based tests such as the Wilcoxon Rank Sum test, for example in [25]. Nonparametric tests are used in our evaluation as our data is not normally distributed. Additionally, a significance level of α=0.05 is used in the evaluation.

VI. Results and Evaluation

A. Real-world Human-robot Collaboration Validation

Fig. 4 visually demonstrates the implementation and validation of the developed approach. In this human-robot collaborative task, the action space is discrete and of size 14. In stage 1, the user demonstrates their sequences to the robot. In stage 2, while the robot is pre-training with demonstration data, the user assembles part of their sequence. Stages 3 and 4 show the robot retrieving the incorrect part, and the user rejecting the selection. Stage 5 shows the robot retrieving the correct and final part of the sequence, which the user accepts and adds to their workspace, completing the task.

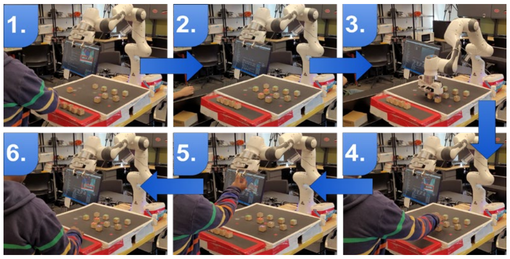


Fig. 4. Implementation and validation of the developed approach in real-world human-robot collaborative tasks.

B. *Overall Efficiency Ratings Analysis and Comparison*

Fig. 5 compares the efficiency ratings between the state-machine-based approach and the developed one from the user study. Visual inspection of the charts shows a more even distribution of responses in the high range for this study's ratings when compared to the state-machine method, which receives more "excellent" ratings. Performing the Brunner-Munzel test produces a *p*-value of 0.47, which is beyond our designated significance level of *α*=0.05. We determine that there exists no statistically significant difference between the datasets. We also consider why efficiency ratings for the deep Q-network-based approach to the task were generally lower than the state-machine counterpart, particularly in the frequency of "excellent" ratings. We presume that this is due to the sample-intensive nature of reinforcement learning that, while most effective over longer periods of time, may appear to be inefficient for short interactions.

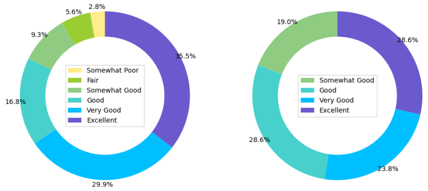


Fig. 5. Comparison of efficiency ratings for state-machine (left) and DQN (right) approaches.

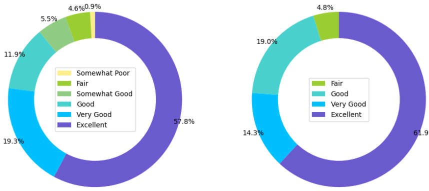


Fig. 6. Comparison of safety ratings for state-machine (left) and DQN (right) approaches.

C. *Safety Ratings Analysis and Comparison*

We compare the two methodologies' ratings on the metric of safety. Fig. 6 shows many similarities between the two approaches, and the Brunner-Munzel test shows no statistically significant difference with a *p*-value of 0.74. But from the number of "excellent" ratings, we find that more participants prefer the deep Q-network-based approach. As noted before, safety is a contributor to human feelings of trust in human-robot interaction and is therefore useful to evaluate.

D. *Comfort Ratings Analysis and Comparison*

The comparison of participants' comfort throughout the interaction with the robot is presented in Fig. 7. First, from the percentage of ratings of "excellent", participants feel more comfortable with the developed approach. In addition, these distributions appear to be similar upon visual inspection and this hypothesis is confirmed with the BM test. The test produces a final p-value of 0.24, which is over the selected significance level. Similar to safety, overall feelings of comfort in a human-robot interactive context assist in better understanding trust.

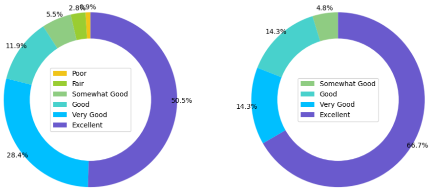


Fig. 7. Comparison of comfort ratings for state-machine (left) and DQN (right) approaches.

E. *Efficiency Subcategories Ratings Analysis*

As previously noted, the metrics (reliability of actions, selected component accuracy, and learning speed) shown in Fig. 8 are factors contributing to the rating of efficiency. As such, they may be used to determine which areas are most impactful. Ratings across these areas appear to be fairly similar, though the accuracy of the robot's selections appears to have been, on average, rated slightly higher. In addition, from the ratings of "Excellent" and "Very Good" by participants, it can be observed that the developed approach of this work offers positive impacts on effective human-robot collaboration in shared tasks.

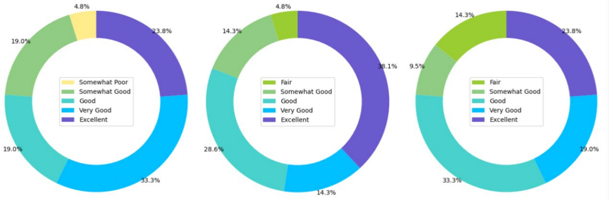


Fig. 8. Participants' ratings on the reliability of actions (left), selected component accuracy (middle), and learning speed (right).

F. *Adaptability Ratings Analysis*

The metric pertaining to the adaptability of the system is also examined. Similar to other ratings by participants, the deep Q-network-based approach to a collaborative assembly task is commonly felt to be considerably adaptable, receiving support at lowest as "somewhat good" up to "excellent," as shown in Fig. 9.

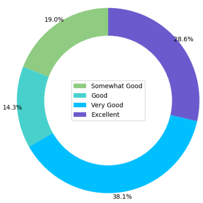


Fig. 9. Adaptability ratings.

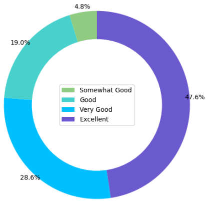


Fig. 10. Trust ratings.

G. *Trust Ratings Analysis*

The metric of trust is valuable to consider in any human-robot interaction. As demonstrated in Fig. 10, with our developed approach, there exists a high level of trust between participants of this study and the DQN-enabled robot that they interact with. In addition, Fig. 11 assists in understanding potential biases towards the metric of trust given the user's personal ratings of previous experience and overall attitude towards robots. As might be expected, participants' attitudes towards robots uncover perhaps a slight bias in their rating of the presented system, with those that have higher levels of interest in robotics rating their trust higher. Additionally, it appears that those with more previous experience with robots tend to rate their feelings of trust lower after interacting with the system.

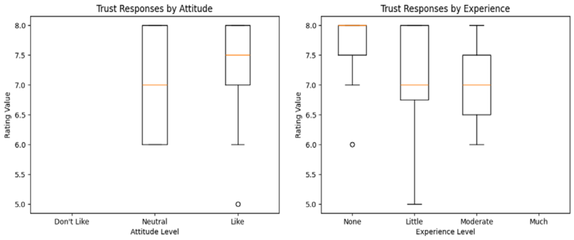


Fig. 11. Participants' trust responses by attitude and experience.

VII. Conclusions and Future Work

In this work, we designed and tested a deep Q-learning-based TLPC implementation for a collaborative robot system in human-robot assembly contexts through a user study. Using the collected data, we compared this approach against a previous state machine-enabled approach to establish potential differences in subjective ratings of the systems for the designated task. Additionally, we examined other metrics as well to gather further insights into users' feelings about the system. We found that, while no statistically significant differences exist between the two approaches for the metrics, the ratings of participants across the board for the developed approach of this work were higher. Participants in the user study indicated that they felt a high degree of safety, comfort, and trust during the interaction with the robot.

While the developed approach is implemented and assessed in this study, assembly tasks in industrial settings are typically more complex than what is being validated. Additionally, given a non-expert demonstrator, the quality of the initial learning phase would certainly be impacted as well. The proposed approach, especially with its property of user feedback, offers the opportunity for the robot to extend its learning beyond its demonstration data, surpassing its initial performance and potentially even the human's. We seek to use the results of this work to consider potential improvements to TLPC as well as alternatives that may perform better in such contexts.

Acknowledgment

This work is supported in part by the National Science Foundation under Grants CMMI-2338767, CMMI-2301678, and CNS-2117308.

References

[1] I-scoop. "Industry 4.0 and the fourth industrial revolution explained." https://www.i-scoop.eu/industry-4-0/ (accessed).

[2] H. Diamantopoulos and W. Wang, "Accommodating and Assisting Human Partners in Human-Robot Collaborative Tasks through Emotion Understanding," in *2021 International Conference on Mechanical and Aerospace Engineering*, 2021, pp. 523-528.

[3] Y. Sun, W. Wang, Y. Chen, and Y. Jia, "Learn How to Assist Humans Through Human Teaching and Robot Learning in Human-Robot Collaborative Assembly," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 52, no. 2, pp. 728-738, 2022.

[4] J. Leng *et al.*, "Industry 5.0: Prospect and retrospect," *Journal of Manufacturing Systems,* vol. 65, pp. 279-295, 2022.

[5] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and Industry 5.0—Inception, conception and perception," *Journal of Manufacturing Systems,* vol. 61, pp. 530-535, 2021.

[6] Y. Lu, J. S. Adrados, S. S. Chand, and L. Wang, "Humans Are Not Machines—Anthropocentric Human–Machine Symbiosis for Ultra-Flexible Smart Manufacturing," *Engineering*, pp. 734-737, 2021.

[7] W. Wang, R. Li, Y. Chen, Z. M. Diekel, and Y. Jia, "Facilitating Human-Robot Collaborative Tasks by Teaching-Learning-Collaboration From Human Demonstrations," *IEEE Transactions on Automation Science and Engineering,* vol. 16, no. 2, pp. 640-653, 2018.

[8] W. Wang, R. Li, Z. M. Diekel, Y. Chen, Z. Zhang, and Y. Jia, "Controlling Object Hand-Over in Human–Robot Collaboration Via Natural Wearable Sensing," *IEEE Transactions on Human-Machine Systems,* vol. 49, no. 1, pp. 59-71, 2019.

[9] S. Joshi, S. Kumra, and F. Sahin, "Robotic grasping using deep reinforcement learning," in *2020 IEEE 16th International Conference on Automation Science and Engineering*, 2020, pp. 1461-1466.

[10] N. M. Gomes, F. N. Martins, J. Lima, and H. Wörtche, "Reinforcement Learning for Collaborative Robots Pick-and-Place Applications: A Case Study," *Automation*, vol. 3, no. 1*,* pp. 223-241.

[11] A. Shafti, J. Tjomsland, W. Dudley, and A. A. Faisal, "Real-World Human-Robot Collaborative Reinforcement Learning," in *2020 IEEE/RSJ IROS*, 2020, pp. 11161-11166.

[12] G. Modery and W. Wang, "Assisting Humans in Human-Robot Collaborative Assembly Contexts through Deep Q-Learning," in *IEEE Undergraduate Research Technology Conference*, 2024 2024, pp. 1-6.

[13] G. Dulac-Arnold, D. Mankowitz, and T. Hester, "Challenges of Real-World Reinforcement Learning," ed, 2019.

[14] O. Obidat, J. Parron, R. Li, J. Rodano, and W. Wang, "Development of a Teaching-Learning-Prediction-Collaboration Model for Human-Robot Collaborative Tasks," in *IEEE International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2023, pp. 728-733.

[15] O. Obidat, G. Modery, W. Wang, X. Guo, and M. Zhou, "A Multifaceted User Study for the Teaching-Learning-Prediction-Collaboration Framework in Human-Robot Collaborative Tasks," in *IEEE International Conference on Automation Science and Engineering*, 2024, pp. 2895-2900.

[16] Y. Li, "Deep Reinforcement Learning," *CoRR,* 2018.

[17] M. Shaili and A. Anuja, "Double Deep Q Network with Huber Reward Function for Cart-Pole Balancing Problem," *International Journal of Performability Engineering,* vol. 18, pp. 644-653, 2022.

[18] H. Hasselt, *et al.*, "Deep Reinforcement Learning with Double Q-Learning," *the AAAI Conference on Artificial Intelligence,* 2016.

[19] N. Akalin and A. Loutfi, "Reinforcement Learning Approaches in Social Robotics," *Sensors*, vol. 21, no. 4*,* doi: 10.3390/s21041292.

[20] T. Hester *et al.*, "Deep Q-learning From Demonstrations," *the AAAI Conference on Artificial Intelligence,* vol. 32, no. 1, 2018.

[21] S. Bier, R. Li, and W. Wang, "A Full-Dimensional Robot Teleoperation Platform," in *2020 IEEE International Conference on Mechanical and Aerospace Engineering*, 2020, pp. 186-191.

[22] C. Hannum, R. Li, and W. Wang, "A Trust-Assist Framework for Human-Robot Co-Carry Tasks," *Robotics,* vol. 12, no. 2, pp. 1-19, 2023.

[23] J. Parron, R. Li, W. Wang, and M. Zhou, "Characterization of Human Trust in Robot through Multimodal Physical and Physiological Biometrics in Human-Robot Partnerships," in *IEEE International Conference on Automation Science and Engineering*, 2024, pp. 1-6.

[24] B. C. Kok and H. Soh, "Trust in Robots: Challenges and Opportunities," *Current Robotics Reports,* vol. 1, pp. 297-309, 2020.

[25] J. D. Karch, "bmtest: A Jamovi Module for Brunner–Munzel's Test—A Robust Alternative to Wilcoxon–Mann–Whitney's Test," *Psych*, vol. 5, no. 2, pp. 386-395, 2023.