Master Thesis

Teaching Robots: Uncovering User Preferences and Interaction Patterns

by

Konstantinos Christofi

(KCI270)

First supervisor: Dr. Kim Baraka
Second reader: Dr. Karine da Silva Miras de Araujo

August 24, 2023

Submitted in partial fulfillment of the requirements for the Master of Science in Artificial Intelligence

Teaching Robots: Uncovering User Preferences and Interaction Patterns

Konstantinos Christofi
Vrije Universiteit Amsterdam
Amsterdam, The Netherlands
k.christofi@student.vu.nl

ABSTRACT

This thesis digs into a deep exploration of human-robot collaboration, specifically focusing on the innate human ability to interact intuitively with robots and provide feedback. The primary objective of this research is to explore the intricate dynamics of human-robot interaction, with a particular focus on the robot learning aspect with the help of human teachers. The study delves into the intricate mechanics of user-driven teaching and interaction patterns, specifically examining the temporal dynamics of user engagement with teaching modalities: Demonstrations and Evaluative Feedback. Moreover, it investigates the potential impact of diverse demographic factors on these interaction patterns, seeking to decipher the nuanced influences that shape human-robot dynamics. Additionally, this research aims to explore the various factors that drive users' interactions with robots, delving into the intricate motivations and preferences that guide their engagement. To achieve these insights, the study employs a learning algorithm that leverages Policy Shaping and Q-learning Reinforcement Learning methods, which operate in the background as users interact with the robot. Through a series of experiments involving 58 participants engaging with the robotic arm, the data we collected involving the teacher's behaviour during training, undergoes meticulous analysis. Notably, while statistical significance is absent between the average time to intervene and demographic variables, a distinct pattern emerges: participants demonstrate a tendency to provide evaluative feedback earlier in their interactions compared to demonstrations. This exploratory analysis lays the foundation for future research attempts, suggesting the potential for more expansive investigations with larger participant groups. Ultimately, this study marks a significant step towards comprehending the complex dynamics of human-robot interaction, offering insights that lead the way for further exploration and understanding of how humans interact with robotic counterparts.

1 INTRODUCTION

In a world where technology advances relentlessly, our interactions with machines often remain mechanical. But what if we could unlock a new era of intuitive and natural communication with robots?

Within this swiftly evolving technological landscape, human-robot interaction (HRI) [32] holds a pivotal position, exemplifying the intricate blend of human ingenuity and artificial intelligence [36]. As technology continues its march, robots are no longer limited to mechanistic roles but are becoming integral collaborators across various sectors. This study embarks on a deep exploration of human-robot collaboration, specifically focusing on the innate human ability to interact intuitively with robots and provide crucial feedback.

This research holds a specific focus on investigating multiple pivotal aspects. A core objective of this study centers around understanding the underlying dynamics of interactions where humans teach robots. This inquiry operates within a theoretical framework grounded in Learning from Demonstration, Learning from Feedback, and Reinforcement Learning algorithms. The fusion of these methodologies underpins the empirical exploration of user interaction preferences and interaction dynamics. Moreover, our exploration aims to dig into the manner in which humans employ demonstration and feedback modalities to instruct a robotic arm within the constraints of a predetermined quota of teaching steps. This scenario necessitates teachers to make strategic decisions about when and how to intervene in the teaching process of the robotic arm. Ultimately, the insights gained from this investigation can provide a foundational framework for the development of oracles - simulated human teachers that evaluate interactive learning algorithms and potentially simulate human behavior when it comes to robot teaching.

The implications of this inquiry extend far beyond technology, shaping the dynamic between humans and machines. Robots, once limited to routine tasks, now have the potential to understand and respond to human intentions, enabling fluid collaboration in sectors like manufacturing [9, 31], healthcare [17, 29], and daily life [27]. Unraveling the intricacies of human-robot interaction and feedback delivery becomes vital for harnessing the full potential of intelligent robots.

The concept of robots learning from human experts holds tremendous potential, offering advantages to both robots and humans. For robots, this approach accelerates their learning curve, allowing them to acquire specialized skills and contextual insights swiftly. This equips them to handle intricate tasks effectively, transforming industries and enhancing efficiency.

For humans, the collaboration with robots that learn from their expertise is equally rewarding. Teaching robots prompts humans to refine their own knowledge, deepening their understanding of the subject matter. This partnership fosters engagement and empowers humans to redirect their efforts towards more creative endeavors. As robots take on routine tasks, humans can focus on innovation, bolstering productivity.

Ultimately, the fusion of human expertise and robotic learning reshapes industries and human-robot interaction, promising a future where machines and humans co-evolve for mutual benefit.

This research seeks to dissect the mechanics of human-robot interaction, focusing on teaching robots and the dynamics governing the symbiotic partnership. Central to this exploration is the question: "Are there consistent patterns in teaching behaviors in user interactions with a robot?" The research aims to explore demographic influences on the dynamic, shedding light on how users naturally interact with robots.

The interconnected objectives guiding this study aim to uncover the dynamics of user-driven teaching and interaction preferences:

(1) How do users' interaction patterns with a robot vary, and what is the timing of teaching modalities - Demonstrations and Feedback - in human-robot interaction?
(2) What influence do demographic variables have on users' preferences for teaching modalities, and how do these variables shape users' interaction patterns?
(3) What are the essential factors that guide users' choices between demonstrations and feedback, and how do these factors impact the dynamics of user-driven interactions?

These objectives collectively aim to unravel the intricate fabric of human-robot teaching interaction, enriching the understanding of user behavior and preferences. The significance of this research spans academia and practical applications, advancing the comprehension of interaction, informing design, and addressing knowledge gaps.

As we delve into this exploration, it is essential to acknowledge the study's scope and limitations. The research offers valuable insights into user preferences but is confined to specific demographic variables and self-reported data. The study's controlled setup may limit generalizability, and potential biases need consideration. The research methodology employs a mixed-methods approach, combining quantitative analysis of interactive experiments with qualitative insights from post-task questionnaires.

The subsequent sections offer a comprehensive exploration of user-driven teaching modalities and interaction dynamics in human-robot interaction. Chapter 2 delves into related work, positioning our study within the broader context. In Chapter 3, we detail the overview of interactive learning methods we use, outlining the integration of Learning from Demonstration, Learning from Feedback and Reinforcement Learning techniques. Chapter 4 presents the study design, including the data collection, and the robot teaching phase followed by the report of the main findings in compact form in Chapter 5. The critical interpretation is reported in Chapter 6, followed by the future research directions and it also reflects on challenges, lessons, and potential improvements. This structured framework systematically unravels the study's progression, culminating in a comprehensive exploration of the research topic.

2 RELATED WORK

In recent years, the domain of teaching robots has witnessed significant advancements as researches strive to enhance the adaptability and efficiency of instruction-following agents. One prevailing challenge in this realm revolves around the limited capacity of existing agents to seamlessly incorporate online natural language supervision and adapt to new instructions swiftly [8]. Moreover, even with ample demonstrations, these agents often struggle to learn even rudimentary policies, highlighting the need for innovative paradigms.

A method for overcoming this challenge is mentioned in [8] and involves a "human-in-the-loop" framework, in which human teacher imparts instructions to the robot, which then autonomously executes actions to fulfill the teacher's goals. The approach encompasses both the resolution of the teacher's goals and the autonomous planning of action sequences within the environment, solely based on provided demonstrations. Notable, even when instructions and clear and explicit, prevailing methods occasionally fail to accomplish the desired tasks effectively. To surmount these limitations, a recent intervention framework has emerged that capitalizes on natural language corrections to predict and rectify errors during execution. These corrections, which come from human evaluation feedback, encompass subtle nuances such as missed grasps or misaligned end-effector placements. The intervention framework leverages these corrections to dynamically adapt the robot's actions in real-time, enabling a symbiotic connection between the human teacher and the robot.

The research yields multifaceted outcomes encompassing both quantitative and qualitative assessments. Qualitative dimensions, including ease of use, control smoothness, and the potential for adopting the control strategy in subsequent tasks, have been examined. Additionally, in order to experimentally assess the efficiency of the framework, the study digs into crucial performance measures including task completion rates.

The study's findings unveil the efficacy of the intervention framework termed "LILAC." Positioned as an extension of LILA, LILAC serves as a robust platform for acquiring natural language interfaces in the domain of human-robot collaboration. A hallmark feature of LILAC lies in its capacity to assimilate natural language corrections in a data-driven manner, contributing to its refined capabilities [8]. LILAC significantly outperforms imitation learning and LILA baselines in critical subtasks like grasping, transfer, and full task completion. These results emphatically validate the veracity of the study's primary hypothesis. A parallel hypothesis asserts that LILAC is preferable than baseline techniques in terms of general usability as determined by respondents' subjective survey replies. Results from the survey back up this claim, showing a clear preference for LILAC in terms of things like "ease of use," "intuitiveness," and "willingness to use again."

In a related line, methods for maximizing human-robot interaction in reinforcement learning scenarios have also been explored in the literature by Yu and Short [37]. The study uses an Active Feedback Learning methodology that is supported by a thorough feedback mechanism that validates the learning process. This learning paradigm is made up of the two fundamental elements of demonstration and prediction. The projection of potential proper and erroneous actions based on the existing situation is included in predictive capabilities. A set of organized experiments serve to realize the study's empirical component. Researchers created an oracle that mimics human evaluations by continuously giving accurate feedback. The experimental protocol encompassed a substantial number of trials, specifically 1000, and results were subjected to averaging to mitigate the influence of random fluctuations. It is noteworthy that, in contrast to the ongoing research, the current study did not undertake the task 1000 times for each policy collected. Nonetheless, this path suggests itself as a promising one for future exploration, serving as a potential center of attention for additional research.

Beyond the scope of specific frameworks, the larger field of interactive reinforcement learning has explored the symbiotic link between agent learning and human evaluative feedback [19]. This option gives agent the chance to learn from both professional and non-experts who might not have a deep understanding of agent design and programming. Within this context, agents ingest human evaluations of their actions, as well as suggestions and directives, enriching the learning process.

The notion of Social Human Robot Interaction (HRI), which refers to interactions between robots and people aided by natural communication forms such as voice, facial expressions and gestures, has also been incorporated into the progression of HRI [19]. This enables humans, even those without extensive training, to interact intuitively with robots, thereby expediting task completion and minimizing the human effort required. As Lin et al. [19] mention, current literature offers a limited scope of empirical experiments that delve into the specific realm of human interaction with robots and their corresponding feedback provision. Nevertheless, the environment demonstrates the identification of various possible study directions despite this restriction. In order to fill this gap, the present study embarks on an exploration of the inherent dynamics underlying human-robot interactions.

Specifically, the investigation aims to decipher the intuitive patterns by which individuals interact with robots and subsequently provide rich feedback. This research endeavors to illuminate the intricate interplay between humans and robots, shedding light on the mechanics of communication and feedback in this evolving domain.

3 OVERVIEW OF INTERACTIVE LEARNING METHOD

Our proposed system implementation is built on the work of previous students of Vrije Universiteit Amsterdam [4]. It employs a number of techniques, including Learning from Demonstration and Reinforcement Learning algorithms such as Policy Shaping (3.1.1) and Q-Learning (3.1.2). Subsection 3.1 describes these methods in depth, whereas subsection 3.3 focuses on their implementation within our proposed system (Figure 1). The implementation can be broadly divided into two modalities: the first involves the human teacher demonstrating actions to the agent, and the second involves the agent utilizing Policy Shaping and Q-Learning to determine optimal actions for specific states. Furthermore, the human teacher can provide evaluative feedback on the agent's state-action pairings. Human teachers can choose either of the modalities they prefer at any point in time. By integrating these techniques, we assume that our system promotes a successful collaboration between the human teacher and the agent, with the assumption of improving the agent's decision-making abilities and overall performance in completing the desired task.

3.1 Background

3.1.1 Policy Shaping. In Reinforcement Learning (RL), a strategy known as Policy Shaping treats human input as an assessment of the course of action rather than traditional rewards. Instead of relying solely on a reward signal to guide the learning agent's behavior, policy shaping leverages human expertise to shape the policy directly. It allows a human teacher to directly affect the behavior of the learning agent by providing positive or negative feedback on specific actions in different states [18]. This feedback encourages or discourages particular actions in particular states which directly affect the agent's action policy. Policy shaping seeks to improve learning and optimize the agent's behavior by interpreting human input as assessments of action choices [6, 13, 16]. Cederborg et al. [6] mention that by incorporating human feedback through policy shaping, the knowledge and skill of the human teacher can be useful to the learning agent. This method makes learning more effective and can potentially accelerate the convergence to optimal behavior.

In this work, we use the policy shaping algorithm that was introduced by Griffith et al. [13]. During its learning process, human teacher has direct access in communicating with the agent. On a particular state, the agent can receive a "good"/"bad" feedback after taking a particular action. As it is also mentioned in the Griffith et al. [13] paper, it is possible that the human providing the feedback, knows any number of different optimal actions in a state. However, we assume that the provided feedback will be based on the best possible action. Yet, the probability of an action α in a certain state s is optimal, is independent of the labels assigned to the other actions. Therefore, the probability (s, α) is optimal and selected amongst other actions in that state, can be calculated by counting the difference between the "good" or "bad" labels associated to it, $\Delta_{s,\alpha}$. The binomial distribution can be used to calculate the probability that (s, α) is optimal as follows:

$$p_c(\alpha|s) = \frac{C^{\Delta_{s,\alpha}}}{C^{\Delta_{s,\alpha}} + (1 - C^{\Delta_{s,\alpha}})} \tag{1}$$

The variables in equation (1) refer to:

- $C^{\Delta_{s,\alpha}}$: The number of "good"/"bad" labels

3.1.2 Q-Learning. The Q-Learning algorithm is a well-known technique in the field of Reinforcement Learning [22, 35]. It is intended to solve Markov Decision Processes (MDPs) [25, 34], which involve decision-making in dynamic and uncertain contexts. Q-learning works by teaching an agent the best policy for making decisions by maximizing the predicted cumulative rewards over time.

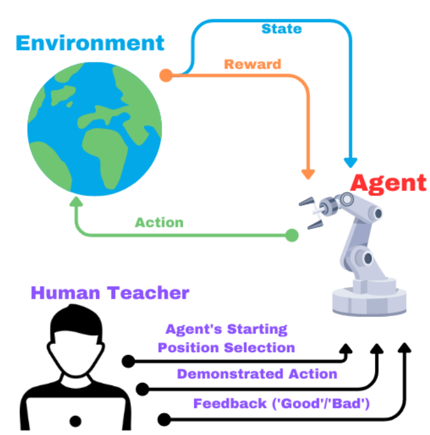


Figure 1: ***Overview of teaching/learning setup***

At its core, Q-learning employs a value function known as the Q-function or Q-value, which represents the expected future rewards that an agent can obtain by doing a specific action in a specific state. The algorithm iteratively adjusts the Q-values based on the agent's experiences in the environment. The agent explores the environment by trial and error, receiving rewards or penalties for its activities and updating its Q-values accordingly. The agent eventually converges on an optimal Q-function that directs its decision-making process, allowing it to select activities that maximize the projected long-term benefits. Q-learning has proven to be an effective and adaptable strategy for training agents in a variety of domains, from gaming to robotics [33, 35].

At the beginning of the training process, the Quality Matrix, often known as Q-Matrix [10], is initialized with dimensions NxZ, where N is the number of the possible states the agent can be in a specific environment and Z is the number of the possible actions that the agent can perform. After the initialization of the Q-Matrix and its values - which in our work all q-values are initialized as 1 - the matrix is updated by using:

$$Q_{new}(s_t, \alpha_t) = Q(s_t, \alpha_t) + \alpha[r_t + \gamma\ max_{\alpha_{t+1}} Q(s_{t+1}, \alpha_{t+1}) - Q(s_t, \alpha_t)] \tag{2}$$

The variables in equation (2) refer to:

- s_t and s_{t+1}: current and next states of the environment.
- α_t and α_{t+1}: current and next actions chosen by the agent.
- α: learning rate, $\alpha \in [0,1]$. It determines the extent to which new information overrides existing knowledge.
- r: reward value after taking a particular action.
- γ: discount factor, $\gamma \in [0,1]$. It quantifies the relative importance of immediate rewards compared to future rewards in the agent's decision-making process.

3.1.3 Learning from Demonstration. Learning from Demonstration (LfD) refers to a RL approach in which a robot or agent learns a task by observing and analyzing the actions performed by a human teacher. LfD surrounds various techniques, including imitation learning [14] and inverse reinforcement learning [23]. In imitation learning, the agent attempts to replicate the expert's behavior directly, while in inverse reinforcement learning, the agent uses the reward function from the expert's demonstrations and learns a strategy to maximize that reward [2, 3].

LfD aims to enable robots or agents to learn new tasks by leveraging the knowledge of human demonstrators. This learning paradigm is especially beneficial when the agent is not executing the same task as the human expert or when the training data is limited [5].

Ravichandar et al. [26] mention that learning from demonstration methods can be categorized based on the technique by which demonstrations will be performed. Despite there are three categories of demonstrations (Kinesthetic Teaching (2a), Teleoperation (2b) and Passive Observation (2c)), in this work we emphasized on Teleoperation which translates the external input to a robot movement.

3.2 Learning Algorithm

The learning algorithm adopted in this study forms a synthesis of the two methodologies we mentioned earlier - Policy Shaping (Section 3.1.1) and Q-Learning (Section 3.1.2). This algorithmic fusion harmonizes the strengths of both approaches, enriching the interaction dynamics between users and the robotic system.

3.2.1 Q-Learning Algorithm. Within the Q-Learning Algorithm, action probabilities are derived from Q-values associated with specific states. This transformation is facilitated through the application of the softmax function, which molds Q-values into a probability distribution. Notably, Q-values signify the anticipated cumulative rewards an agent can accrue by executing distinct actions from a given state. By normalizing Q-values through softmax, probabilities are assigned to each action, reflecting the likelihood of selection. This technique capitalizes on the relative differences between Q-values, allocating higher probabilities to actions with elevated Q-values while ensuring a balanced distribution across all potential actions.

3.2.2 Policy Shaping Probability Calculation. In the realm of Policy Shaping, the probability for each action is calculated using Equation 1. When feedback is available for a state, the algorithm computes action probabilities based on feedback magnitude. The negative feedback values are transformed into a bounded range spanning from -50 to 0. This transformation mitigates the disproportionate influence of extreme negative feedback on probability computation. By performing this conversion, the algorithm associates negative feedback values with probabilities that guide the agent away from actions linked with substantial negative feedback, while still facilitating exploration.

3.2.3 Probabilistic Synthesis and Action Selection. Subsequently, the probabilities indicating an action's selection by the agent, derived from both Policy Shaping and Q-Learning algorithms, are amalgamated. This fusion involves weighting and summing the probabilities to yield a comprehensive list of probabilities for all actions corresponding to a specific state. The ensuing step entails the stochastic selection of an action from this list, executed with probabilities that were pre-distributed. This stochastic action selection mechanism ensures a dynamic, data-driven approach to decision-making within the robotic system.

In summation, the learning algorithm deployed in this study merges Policy Shaping and Q-Learning, resulting in a harmonious interplay of methods. This algorithmic synergy not only enhances the robotic system's adaptability but also augments the user experience by enabling nuanced interactions aligned with both human preferences and algorithmic optimization.

3.3 Implementation

For the implementation of this project, we designed and implemented a graphical user interface (GUI) (3) to facilitate the interaction between a human teacher and an Agent operating in the OpenAI Gym environment for the Franka Emika Panda robot [11]. The interaction between the human teacher and the agent consists of two interconeccted phases that are stated below.

During the initial phase, the human teacher has control of the agent and uses the keyboard to illustrate the desired actions for specific states. Subsequently, the agent is rewarded based on the environment and the task at hand. This reward serves as feedback for the agent's performance, guiding its learning process.

(a) *Kinesthetic teaching* (b) *Teleoperation* (c) *Passive observation*

Ravichandar H, et al. 2020.
Annu. Rev. Control Robot. Auton. Syst. 3:297–330

Figure 2: ***Examples of the three categories of robot demonstrations [26].***

In the second phase, the agent chooses actions independently using a combination of Policy Shaping (3.1.1) and Q-Learning (3.1.2) approaches that are mentioned above. The Q-Learning algorithm updates the Q-Matrix, allowing the agent to estimate the expected rewards for different state-action pairs. Furthermore, the human teacher can provide "Good" or "Bad" feedback to the agent, which is used to change the appropriate values in the Policy Shaping table. This feedback enables the agent to fine-tune its policy and improve its decision-making capabilities within the constraints of the environment.

By integrating these interactive components into the GUI, the project enables effective communication and collaboration between the human teacher and the agent, facilitating learning and improving the agent's effectiveness in completing the intended task.

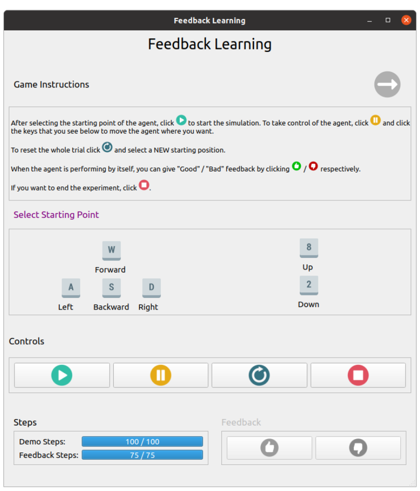


Figure 3: ***Graphical User Interface for Human-Robot interaction***

4 STUDY DESIGN

In pursuit of a comprehensive evaluation of our research, a multifaceted approach was embraced. Firstly, a pilot study was conducted in which the learning algorithm was not a priority, encompassing a group of 29 participants. Within this group of people, a demographic distribution was observed, encompassing 17 females and 12 males, spanning ages from 22 to 51 years. The pilot study aimed to discern the impact of diverse demographic attributes, such as educational background, experience with programming, video game experience etc., on the interaction dynamics with the robotic arm. This first study provided foundational insights to inform subsequent phases of inquiry.

Building upon the insights we gathered from the pilot study, a follow-up study was performed to delve into how teachers can affect robot learning. In this subsequent phase, a group of 30 participants took part, consisting of 21 males and 9 females, with ages spanning from 22 to 59 years. The cumulative findings from both studies compose a complete picture to guide our research.

Our research endeavor comprises a divided structure. The initial phase entails participants interacting with a robotic arm to impart knowledge and thereby instruct it in new tasks. Following the execution of these tasks, participants move to the second phase of the study. This phase entails participants responding to a structured questionnaire, aimed at gathering insights into their overall experiential encounter. This holistic framework not only embraces diverse demographics attributes but also encapsulates the intricate dynamics between teacher input and the learning trajectory of the robotic arm.

4.1 Robot Teaching

After finishing with the initial phase of the experiment involving interaction with the robotic arm, participants are presented with a comprehensive overview of the subsequent experimental process. As the experiment commences, participants are instructed to make strategic selections for the start positioning of the robotic arm. This strategic placement is intended to optimize the robot's learning trajectory throughout the instructional phase.

Following the establishment of the robot's starting position, participants are empowered to start the robotic arm's movement within the designated environment, thereby facilitating task execution. Throughout this phase, participants retain the agency to interact with the robot in two distinct modalities. Firstly, participants can provide a sequence of actions known as a "Demonstration", which the robot imitates. Secondly, participants can deliver evaluative feedback, labeled as "Good" or "Bad", to guide the robot's learning trajectory.

It is explicitly communicated to the participants that their allocations for both demonstrations and feedback are constrained, necessitating strategic selections to optimize the instructional process. To help participant understanding of these concepts, an initial familiarization phase with the robot is undertaken, enabling them to grasp the mechanics of Demonstrations and Feedback. Participants then move into the main instructional portion, when they are given a variety of tasks to do, each of which calls for them to teach the robot a certain skill.

Regarding the Demonstrations and Feedback steps, a rigorous evaluation served as the foundation for the strategy used to calculate the appropriate quantities. In particular, a method was used that involved running a "perfect" policy and a random policy across a span of 1000 episodes. The objective was to discern the point at which the agent learned to complete the task, as measured by the number of steps taken.

After numerous interactions with the robotic arm, in the capacity of an expert I provided a policy to the agent. This policy was founded upon the best feasible actions to be executed within specific states. Subsequent to the creation of these policies, a comparative analysis ensued, resulting in the determination of optimal quantities for Demonstration and Feedback steps. Specifically, the chosen quantum for Demonstrations was established at 100 steps, while Feedback was calibrated to 75 steps.

It should be noted that each task iteration displayed a 3 minute time limit or the culmination of 5 sequential task completions (episodes). The total number of demonstration and feedback steps correspond to the numbers that are mentioned above, 100 steps for demonstration and 75 for feedback.

4.1.1 Data Collection. During the user-robot interactions, an automated data collection mechanism operates in the background [24]. This mechanism captures pertinent information, including states visited by the robot and the corresponding actions undertaken within those states. Additionally, immediate rewards and Q-values associated with these state-action pairs are systematically logged. Moreover, the precise timing of user actions, such as demonstrations or evaluative feedback is precisely recorded. The synergy of these automated data collection components contributes to an extensive repository of insights, paramount for the subsequent analysis and refinement of the instructional process.

4.1.2 Tasks. Leveraging the capabilities of the Panda-Gym environment [11], we harnessed the facility to create customized environments along with bespoke tasks. For the purpose of this thesis, we undertook the creation of two novel tasks, building upon the foundational framework of the pre-existing "Reach" task.

In the first task presented to users, the objective involves guiding the robot towards a designated green sphere, as visually depicted in Figure 4a. When the agent executes actions resulting in the robot deviating from the trajectory leading to the green sphere, an incurred penalty of -1 is applied. Upon successful attainment of the green sphere, an instantaneous reward of 0 is conferred.

Transitioning to the second task, this entails the introduction of both green and red spheres, as illustrated in figure 4b. Participants are asked to teach the robot with steering away from the red sphere, while concurrently steering it toward the green counterpart. Consistent with the initial task, deviations from the trajectory leading to the green sphere incur a penalty. Intriguingly, any interaction with the red sphere incurs a more substantial penalty of -1.5, accentuating the significance of circumventing it.

The third and final task within this paradigm entails a sequential trajectory encompassing both blue and green spheres, ad delineated in figure 4c. The prescribed objective necessitates first reaching the blue sphere and subsequently proceeding to the green sphere. This sequential nature of the task, accentuates the cognitive challenge presented to participants within the instructional framework.

4.2 Questionnaire

The first segment of the questionnaire 9.1 encompassed inquiries aimed at eliciting participants' holistic impressions and perceptions arising from their interactions with the robotic arm. In the context, participants were invited to share insights regarding the instances prompting their decision to provide demonstrations or evaluative feedback. Additionally, participants were queried about the perceived intuitiveness of these mechanisms. Furthermore, participants were encouraged to articulate their strategic rationale underpinning the selection of the agent's starting position, a facet anticipated to yield valuable insights for subsequent data analysis.

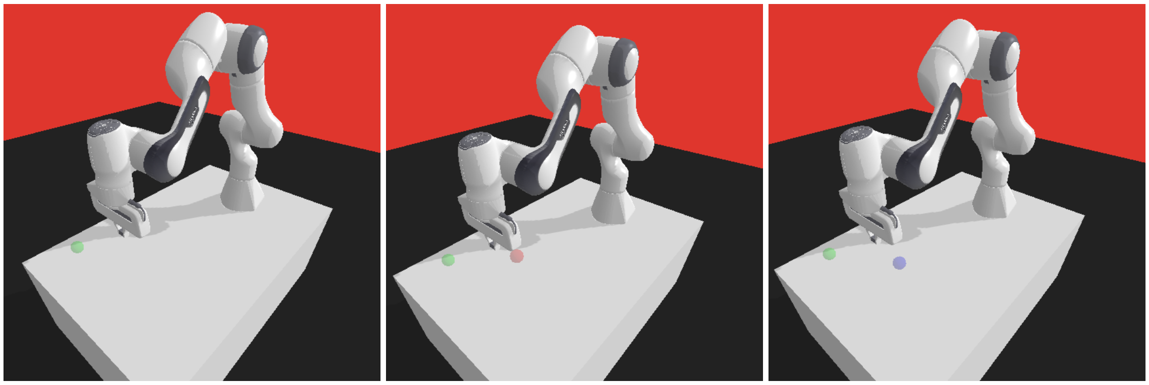
The subsequent component of the questionnaire delved into inquiries that gauged participants' preferences regarding their proclivity to either instruct a robot or employ a pre-trained counterpart, assuming ownership thereof. This line of questioning aimed to illuminate participants inclinations in terms of interaction in imparting knowledge to machines versus utilizing pre-established automated systems.

The final tier of the questionnaire encompassed demographic inquiries, strategically integrated to furnish contextual information for the ensuing data analysis. The demographic inquiries that we collect are gender, age, educational level, pet ownership, teaching/training involvement in profession, programming experience and video games experience. These demographic inquiries encompassed thereby enhancing the comprehensive understanding of participant profiles and backgrounds. Subsequent to the completion of all questionnaire sections, a well-rounded perspective is anticipated to emerge, rendering the dataset enriched and amendable to nuanced analysis.

4.3 Analytical Framework

In the pursuit of comprehensive exploration, our research methodology adopts a mixed-methods approach that harmoniously blends qualitative and quantitative techniques. By embracing both quantitative data collection and qualitative insights, we achieve a holistic understanding of user preferences, interaction patterns, and timing.

Upon completion of the data collection phase, our investigation will transition into an exploratory study aimed at uncovering patterns withing teaching modalities and addressing the core research inquiries. A focal aspect of this analysis involves calculating the temporal disparity between the average time taken by each individual participant to provide demonstrations versus feedback. This calculated difference, denoted as 'delta' (3), plays a pivotal role in shedding light on trends. A positive delta signifies that participants, on average, tend to supply evaluative feedback before



(a) *Reach green sphere* (b) *Task 2 - Avoid red sphere and reach green sphere* (c) *Task 3 - Reach blue sphere first and then green sphere*

Figure 4: ***Example of the three tasks***

demonstrations. Conversely, a negative delta indicates that specific participants tend to present demonstrations prior to offering feedback.

$$\delta = \bar{time}_{demonstration} - \bar{time}_{feedback} \tag{3}$$

Before looking at the average deltas, we will carefully study these values, along with various demographic variables of the people involved to unearth potential patterns and correlations. By exploring into how demographic attributes might influence the selection of teaching modalities, this preliminary analysis promises to offer a deeper understanding of the process. Subsequently, the aggregation of these individual deltas into averaged values will provide insight into participants' actions. This multi-faced approach aims to identify patters of user interaction and their interrelation with teaching modalities, thereby enriching our comprehension of the human-robot interaction. Before Furthermore, an additional aspect of our investigation involves examining how users interacted with the robotic arm through the application of both demonstration and feedback. This examination will entail calculating the average number of demonstration and feedback instances contributed by each participant.

5 RESULTS

To address our primary research inquiry - "What is the relationship between the time of a user's interaction with a robot and the provision of demonstrations versus evaluative feedback in human-robot interaction?" - we harnessed both datasets, encompassing all participants, as we selectively employed parameters pertinent to participant interaction with the robot, whether through demonstration or evaluative feedback. By focusing solely on relevant parameters, we ensured a comprehensive analysis of interaction dynamics.

5.1 Timing patterns in teaching modalities

Initially, we computed the average teaching modality that participants adopted first and the subsequent average time it took for them to employ the alternate method. It is important to note that a subset of participants did not contribute evaluative feedback to the robot, necessitating their exclusion from the sample.The calculated mean overall difference, amounting to 4.63 seconds, unveils a noteworthy trend: participants tend to supply evaluative feedback earlier than demonstrations by a margin of 4.63 seconds. Furthermore, we ventured into a comparative examination of the delta differences with respect to various demographic variables, An intriguing observation emerged from the analysis of delta in relation to gender (Figure 5). The outcomes highlighted a divergence between male and female participants. Specifically, male participants exhibited a mean delta of 7.4 seconds (±30), while their female counterparts displayed a mean delta of 1.3 seconds (±30.5). This indicates that both male and female participants favored supplying feedback ahead of demonstrations. Notably, male participants appeared to take more time to transition from providing evaluative feedback to presenting demonstrations.

In the graph shown in Figure (5), we observe the contrast in average time taken to offer demonstrations and feedback, categorized by gender. The data indicates that both males and females tend to give evaluative feedback before providing demonstrations. Although the middle value between the genders is roughly the same, females exhibit a wider range of differences in the timing of their provided demonstrations and feedback.

To determine the potential statistical significance between gender and delta differences, we executed a two-sided Rank-Sum test [28] with the resulting p-value of 0.42 which means there is no statistical significance between the two groups. Despite the absence of a statistical significance, these insights collectively illuminate the intricate interplay of gender, teaching modalities, and interaction timing within the realm of human-robot interaction.

Additionally, the mean time taken by each participant for their initial interaction, whether it involved providing a Demonstration or Feedback, was calculated. The computed mean time to interact was found to be 35.46 seconds (±17.44). A further analysis involved the comparison of this mean interaction time with the demographic variables collected from participants. Intriguingly, a statistically significant difference in the mean time to interact was observed in relation to gender after executing a two-sided Rank-Sum test [28],

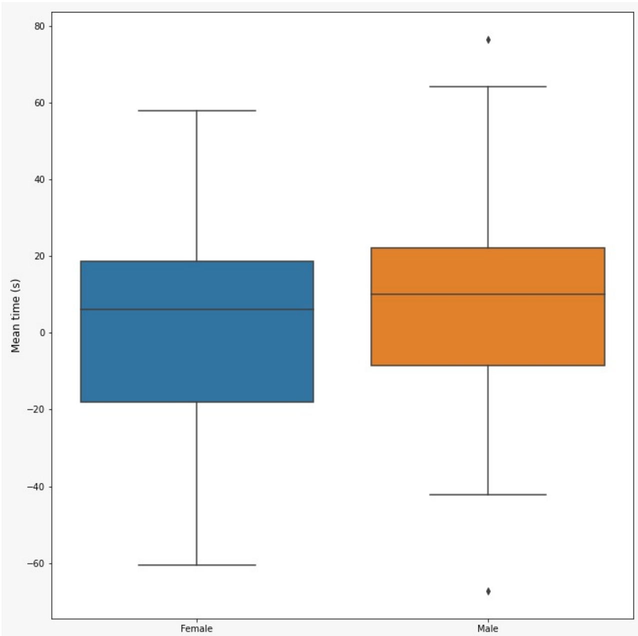


Figure 5: ***Mean difference of average interaction with demonstration versus feedback: The figure illustrates the comparison of average time intervals for delivering demonstrations and feedback, segmented by gender. The data reveals that both males and females generally prioritize providing evaluative feedback over demonstrations. The median time difference between the genders is relatively similar.***

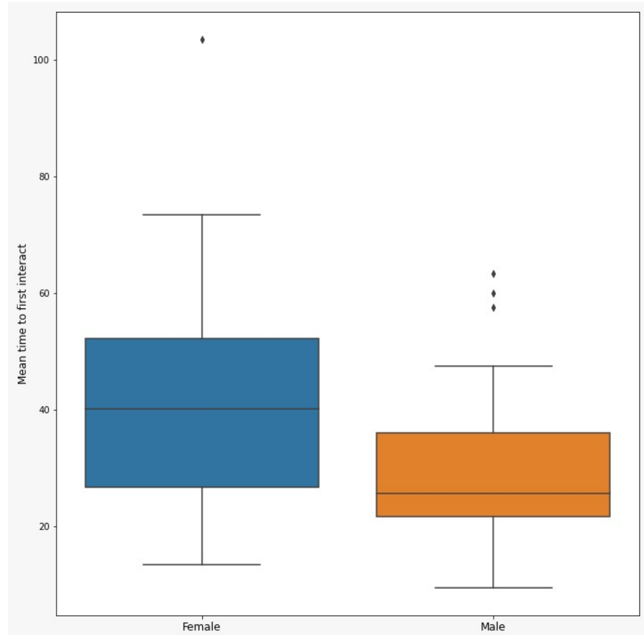


Figure 6: ***Mean time to first interact by gender: The disparity in average time for participants' initial interactions with the robotic arm is evident, categorized by gender. The median duration for male participants proves shorter than that for females, suggesting males tend to engage with the robot earlier than females.***

with a p-value of 0.0057. Specifically, female participants exhibited a mean time to interact of 42.62 seconds (±20), whereas their male counterparts exhibited a mean time of 29.81 seconds (±12.8). This disparity suggests an interesting gender-based variation in the temporal aspects of user interaction with the robotic arm.

In the following figure 6, we observe the contrast in average time for participants' initial interaction with the robotic arm, categorized by gender. The median duration for male participants is shorter than that for females, indicating that males tend to engage with the robot earlier than females. Among male participants, the average interaction times exhibit less variation compared to the more dispersed average times observed among females.

5.2 Differences in quotas usage

In addition to this analysis, we explored potential gender-based differences in interaction patterns with the robot using both teaching modalities that can also be seen in figure (7). We computed the average number of demonstration and feedback steps provided by each participant and conducted the Rank-Sum test [28] once again to ascertain whether any statistically significant differences existed between gender and the provision of demonstration and feedback steps. The mean quantity of demonstration steps employed by females was 60.8 (±26.9), while males averaged 58.9 (±23.1) steps. For feedback steps, females exhibited a mean of 17.3 (±19.8) steps, compared to males with a mean of 22.8 (±18.4) steps. The Rank-Sum test outcomes indicated no statistical significance between gender and the usage of demonstration and feedback steps, yielding p-values of 0.74 and 0.14 respectively.

The box plot displayed in Figure 7 illustrates the distribution of interaction frequencies associated with the teaching modalities "Demonstrations" and "Feedback" among various gender groups. The data indicates that both male and female participants tend to utilize feedback less frequently than demonstrations. Interestingly, the medians for both genders in both teaching modalities are quite similar. However, it's noteworthy that the variability in the number of demonstrations employed is higher compared to the number of feedback instances.

5.2.1 User-Driven Teaching Factors. Furthermore, we conducted an insightful exploration into participants' interactions with the robotic arm by analyzing responses to the open-ended survey questions regarding the situations they chose to provide Demonstrations or Evaluative Feedback. In order to elucidate the patterns underlying participants' decision-making processes, we employed a clustering approach with the assistance of ChatGPT [1]. This involved grouping participants' responses based on recurring themes, such as "Distance to target" or "Robot not following instructions." "Proximity to Goal or Target" includes responses from the questionnaire like "When the robot was away from the target", "Performance" cluster includes answers like "Robot staying for a long time at the same position". "Following Instructions" includes answers "When

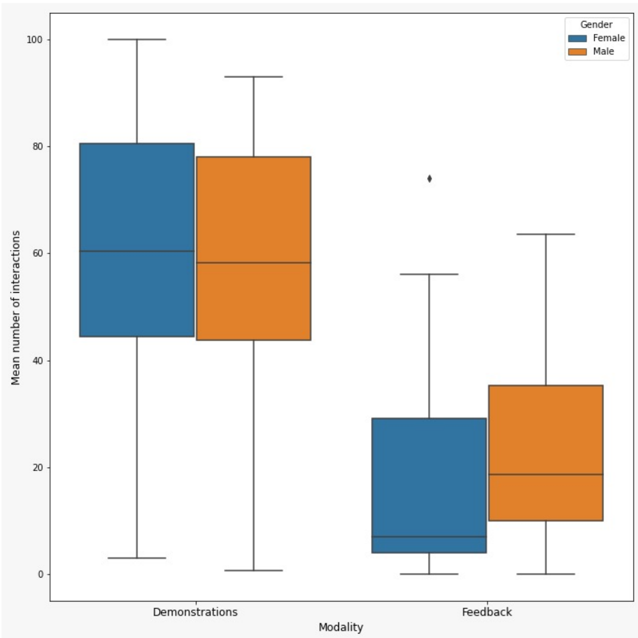


Figure 7: *Gender-Based Interaction Modality: The boxplot graphically represents the distribution of interaction frequencies for two distinct modalities ("Demonstrations" and "Feedback") across different genders. The analysis sheds light on how interaction patterns vary based on both interaction modality and the respondent's gender.*

I wanted to teach a trajectory to the robot" and "Not Following Instructions" cluster includes answers like "When the robot was not following a trajectory that was shown on an earlier demonstration". The outcomes of this clustering are illustrated in Figures 8 & 9, providing an illuminating overview of the situations prompting participants to interact with the robotic arm.

Figure 8 delineates the scenarios in which participants opted to provide Demonstrations. Notably, out of the 59 participants, 39 chose to provide demonstrations in response to the consideration of "Proximity to Goal or Target". Additionally, 11 participants indicated instances where the robot's performance prompted them to provide demonstrations, while 6 participants offered guidance to the robot to ensure task completion. It is noteworthy that 1 participant refrained from providing demonstrations and 2 responses were classified as miscellaneous. Meanwhile, Figure 9 illustrates the situations in which participants chose to offer Evaluative Feedback. The chart indicates that 37 participants provided feedback based on "Proximity to Goal or Target", highlighting the significance of this factor in interaction dynamics. Furthermore, 11 participants opted for evaluative feedback when the robotic arm deviated from the path that the user showed on a previous stage, and 5 participants utilized this feedback modality when the robot's performance did not align with their intentions, following a series of moves that did not guide the robot toward the target. Intriguingly, 3 participants abstained from providing evaluative feedback, while 3 responses were categorized as miscellaneous in nature. This analysis provides valuable insights into the contextual triggers that prompt users to select specific teaching modalities during their interactions with the robotic arm.

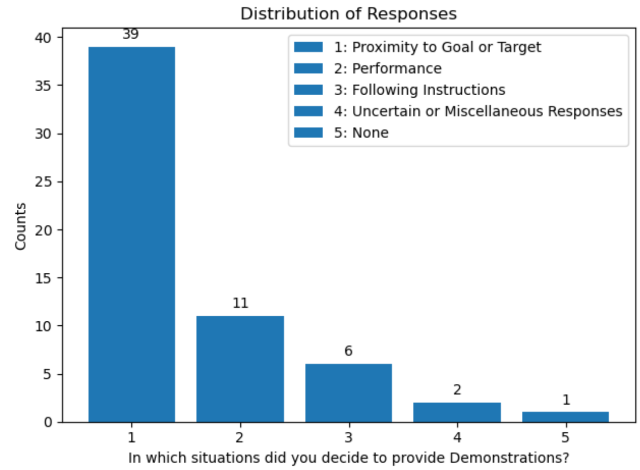


Figure 8: *Situations that participants provided demonstration*

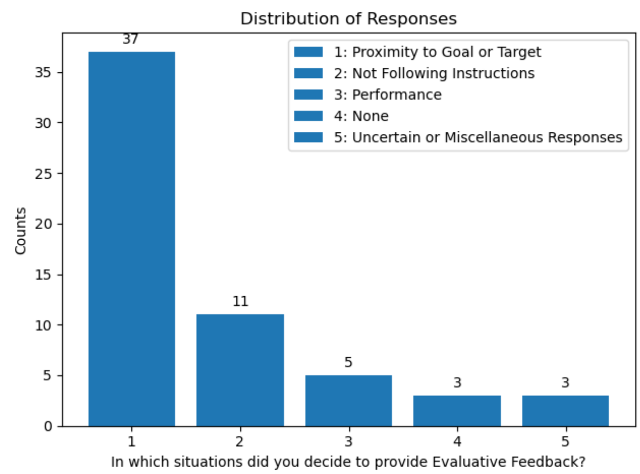


Figure 9: *Situations that participants provided evaluative feedback*

6 DISCUSSION

In pursuit of a comprehensive understanding of human-robot interaction dynamics, this study delved into the intricate mechanics of teaching robots, which underpin the symbiotic partnership between humans and machines. Central to this exploration was the pivotal inquiry: "Are there consistent patterns in teaching behaviors in user interactions with a robot?" The research not only aimed to elucidate these fundamental interaction patterns but also sought to investigate the potential influences of demographic factors on this dynamic interplay, shedding light on the inherent inclinations of users as they naturally engage with robots.

The intertwined objectives that steered this research were crafted to uncover the nuanced dynamics of user-driven teaching preferences and interaction patterns and answer the following questions:

(1) How do users' interaction patterns with a robot vary, and what is the timing of teaching modalities - Demonstrations and Feedback - in human-robot interaction?
(2) What influence do demographic variables have on users' preferences for teaching modalities, and how do these variables shape users' engagement patterns?
(3) What are the essential factors that guide users' choices between demonstrations and feedback, and how do these factors impact the dynamics of user-driven interactions?

Together, these objectives constituted a comprehensive framework that facilitated an in-depth analysis of the complex fabric of human-robot interaction, encompassing engagement timing, demographic influences, and user-driven interaction preferences. Through these endeavors, the study aimed to offer a nuanced comprehension of the intricate symbiosis between humans and robots, fostering a deeper understanding of collaborative dynamics in the realm of human-robot interaction.

The findings obtained from the results section (Section 5) unveil significant insights into the temporal dynamics of user interactions. The analysis, encompassing the mean overall difference between the provision of demonstrations and feedback, indicates that, on average, participants favored supplying evaluative feedback earlier than demonstrations, with a time difference of 4.63 seconds. While the examination of this delta difference alongside diverse demographic factors did not yield statistically significant results, it did illuminate a pronounced divergence between male and female participants. Specifically, male participants exhibited a mean delta of 7.4 seconds (±30), while their female counterparts demonstrated a mean delta of 1.3 seconds (±30.5).

This observed gender-based variation aligns with existing literature suggesting that women tend to provide evaluative feedback earlier than men, driven by several factors:

(1) Perception of Feedback: Women often perceive evaluative feedback, especially negative feedback, as a more informative reflection of their abilities than men [15, 30]. This perception can prompt women to place greater importance on feedback and use it to make quicker adjustments.
(2) Alignment with Peer Feedback: Women display a tendency to align their self-awareness with peer feedback at a faster rate compared to men [21]. Their heightened sensitivity to social cues facilitates a quicker integration of peer feedback into their self-perception.
(3) Self-Image Rationalization: Men may exhibit a tendency to rationalize and maintain an inflated self-image over time, even when confronted with feedback [21]. This behavior might make them less inclined to modify their self-perception based on received feedback.
(4) Emotional Fallout: Men and women might navigate the emotional consequences of feedback differently [7, 21]. Women's heightened emotional sensitivity could lead to quicker adjustments in self-perception following received feedback.

In addition to this investigation, we assessed the mean time participants required for their initial interaction with the robotic arm. The computed mean time to interact was 35.46 seconds (±17.44). Upon scrutinizing this interaction time in relation to demographic variables, a statistically significant gender-based difference emerged with a p-value of 0.0057. Female participants exhibited a mean interaction time of 42.62 seconds (±20), whereas their male counterparts displayed a mean time of 29.81 seconds (±12.8). This disparity underscores an intriguing gender-associated variation in the temporal aspects of user engagement with the robotic arm.

In addition to the previously discussed analysis, we also explored the connection between user interactions within each state and the corresponding updated q-values. The aim of this investigation was to determine if different teaching modalities had a relationship with q-values in specific state-action pairs. The results of this analysis revealed a clear pattern: higher q-values were associated with a greater likelihood of demonstrations being chosen as interactions within those states, and the feedback received in those states tended to be equal or more positive than negative. Conversely, lower q-values were linked to a higher probability of interactions being feedback within those states, particularly negative feedback. It's important to note that when states had an equal number of demonstration and feedback steps, along with an equal number of good and bad feedback steps, there is a high probability that no interactions occurred in those states. However, this analysis may not lead to direct conclusions. Demonstrations can directly impact q-values by indicating successful task completion patterns in teaching behaviors. Conversely, teacher feedback doesn't directly influence q-values; rather, it influences the values in the policy shaping table. These combined factors, along with q-values, affect the probabilities of action selection in specific states. And for that reason, we excluded this analysis from the main findings of the research.

We also attempted an analysis to predict teachers' interactions using machine learning algorithms [20]. This involved training multiple algorithms using the first three episodes of each task during the training phase. The goal was to predict the type of interaction in the final two episodes in a multi-class classification scenario [12]. The three interaction classes were labeled as "None," "Demonstration," and "Feedback." The input variables for the prediction included various robot-related factors such as "State," "Action," "Elapsed-time," "Q-value," and more. Despite achieving a high accuracy rate of 0.85, the analysis did not yield optimal outcomes. The primary issue was that a significant portion of the samples were predicted to fall under the "None" interaction class. It's possible that altering the parameters could lead to improved results, but it's worth considering that predicting teacher interactions might not be reducible to a linear or straightforward decision-based solution.

Certainly, our research journey encounters certain limitations that may have impacted the seamless progression of our study. As we delved into the phase of analyzing data, our intention was to uncover intricate connections and associations among the various data points stemming from participants' interactions with the robot and their responses to the subsequent questionnaire. Regrettably, despite our dedicated efforts, we could not pinpoint significant correlations within the data, potentially owing to the relatively small number of participants in our study. Additionally, the design and implementation of the questionnaire lacked expert oversight, potentially leading to some subjectivity in participants' answers to specific queries. It's worth considering that a more professionally structured questionnaire might have elicited more objective perspectives from participants.

Another point of consideration is the absence of an initial questionnaire prior to the experiment. Such a preliminary survey could have illuminated diverse psychological factors capable of shaping teachers' behaviors during interactions with the robot. With these insights in mind, future research endeavors could adopt a more meticulous approach, aimed at delving into and clarifying additional factors that potentially influence teachers' actions within the context of human-robot interaction. By addressing these limitations and refining our research methodology, a more comprehensive exploration could unfold, providing deeper insights into the intricate web of factors that govern the dynamics between human instructors and robotic counterparts.

7 CONCLUSION

The present study aimed to explore user interactions with robots using demonstration and evaluative feedback, delving into interaction patterns and the timing of teaching modalities. Additionally, we sought to examine the influence of demographic factors on users' preferred teaching modalities and the underlying user-driven teaching factors.

Our analysis unveiled intriguing insights into the temporal dynamics of user interactions. Specifically, participants demonstrated a tendency to provide evaluative feedback earlier than demonstrations, indicating an average time difference of 4.63 seconds. While demographic correlations did not yield significant results, a gender-based discrepancy emerged. Male participants exhibited a mean delta of 7.4 seconds (±30), contrasting with their female counterparts' mean delta of 1.3 seconds (±30.5), aligning with established gender-related feedback behavior trends.

Furthermore, we examined the mean interaction time with the robotic arm, resulting in an average of 35.46 seconds (±17.44). A gender-related disparity was statistically significant, with a p-value of 0.0057. Notably, female participants spent an average of 42.62 seconds (±20), while male participants engaged for an average of 29.81 seconds (±12.8), revealing a gender-specific temporal divergence in user engagement with the robotic arm.

In the realm of user-driven teaching factors, our findings indicated that a majority of participants (over 60%) demonstrated a preference for interacting with the robotic arm based on its position relative to the goal.

Looking ahead, several promising avenues for future research in the domain of human-robot interaction stand ready for exploration. First, the creation and utilization of oracles, which simulate human teaching behavior, could emerge as a potent tool for evaluating interactive learning algorithms. These oracles would enable researchers to comprehensively assess the effectiveness of various learning strategies in a controlled environment, ultimately enhancing the development of more efficient and adaptable robotic learning mechanisms.

Second, the simulation of human behavior in robot teaching scenarios offers a captivating avenue for investigation. By replicating and analyzing the diverse ways in which humans interact with robots, researchers could construct predictive models that guide the design of intelligent robots capable of understanding and responding to human teaching cues. Such models could pave the way for the creation of robots that seamlessly integrate with human workflows, enhancing overall efficiency and collaboration.

In the journey toward deeper human-robot collaboration, the fusion of human expertise and machine capabilities holds the key to unlocking a future where intelligent robots seamlessly augment and amplify human potential.

8 ACKNOWLEDGEMENTS

I am profoundly grateful for the invaluable guidance, support, and insights provided by my dedicated supervisor, Dr. Kim Baraka, whose expertise has been instrumental in shaping this research. I would also like to extend my gratitude to Raj Bhalwankar and Mehul Varma, whose pioneering work has formed the foundation for the learning algorithm at the core of this study. Their contributions have paved the way for the advancements and discoveries presented within this thesis.

REFERENCES

[1] [n.d.]. OpenAI. (2023). ChatGPT (August 3 version) [Large language model]. https://chat.openai.com/
[2] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and autonomous systems* 57, 5 (2009), 469–483.
[3] Christopher G Atkeson and Stefan Schaal. 1997. Robot learning from demonstration. In *ICML*, Vol. 97. Citeseer, 12–20.
[4] Raj Bhalwankar, Mehul Verma, Kim Baraka, and Daniel Preciado. [n.d.]. "What Makes a Good Teacher?" Inter-Individual Differences in Humans who Teach Agents. https://drive.google.com/file/d/1raKnoXFCmp7JihtuPRoUrzGh8BjTezuu/view
[5] Aude Billard, Sylvain Calinon, Ruediger Dillmann, and Stefan Schaal. 2008. Robot programming by demonstration. In *Springer handbook of robotics*. Springer, 1371–1394.
[6] Thomas Cederborg, Ishaan Grover, Charles L Isbell Jr, and Andrea Lockerd Thomaz. 2015. Policy Shaping with Human Teachers.. In *IJCAI*. 3366–3372.
[7] Xuhai Chen, Hang Yuan, Tingting Zheng, Yingchao Chang, and Yangmei Luo. 2018. Females are more sensitive to opponent's emotional feedback: evidence from event-related potentials. *Frontiers in Human Neuroscience* 12 (2018), 275.
[8] Yuchen Cui, Siddharth Karamcheti, Raj Palleti, Nidhya Shivakumar, Percy Liang, and Dorsa Sadigh. 2023. No, to the Right: Online Language Corrections for Robotic Manipulation via Shared Autonomy. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 93–101.
[9] Ilesanmi Daniyan, Khumbulani Mpofu, Boitumelo Ramatsetse, and Adefemi Adeodu. 2020. Design and simulation of a robotic arm for manufacturing operations in the railcar industry. *Procedia Manufacturing* 51 (2020), 67–72.
[10] Jimmy de la Torre and Chia-Yi Chiu. 2016. A general method of empirical Q-matrix validation. *Psychometrika* 81 (2016), 253–273.
[11] Quentin Gallouédec, Nicolas Cazin, Emmanuel Dellandréa, and Liming Chen. 2021. panda-gym: Open-Source Goal-Conditioned Environments for Robotic Learning. *4th Robot Learning Workshop: Self-Supervised and Lifelong Learning at NeurIPS* (2021).
[12] Margherita Grandini, Enrico Bagli, and Giorgio Visani. 2020. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756* (2020).
[13] Shane Griffith, Kaushik Subramanian, Jonathan Scholz, Charles L Isbell, and Andrea L Thomaz. 2013. Policy shaping: Integrating human feedback with reinforcement learning. *Advances in neural information processing systems* 26 (2013).
[14] Ahmed Hussein, Mohamed Medhat Gaber, Eyad Elyan, and Chrisina Jayne. 2017. Imitation learning: A survey of learning methods. *ACM Computing Surveys (CSUR)* 50, 2 (2017), 1–35.
[15] Maria Johnson and Vicki S Helgeson. 2002. Sex differences in response to evaluative feedback: A field study. *Psychology of Women Quarterly* 26, 3 (2002), 242–251.
[16] Taylor Kessler Faulkner, Reymundo A Gutierrez, Elaine Schaertl Short, Guy Hoffman, and Andrea L Thomaz. 2019. Active attention-modified policy shaping: socially interactive agents track. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 728–736.
[17] Maria Kyrarini, Fotios Lygerakis, Akilesh Rajavenkatanarayanan, Christos Sevastopoulos, Harish Ram Nambiappan, Kodur Krishna Chaitanya, Ashwin Ramesh

Babu, Joanne Mathew, and Fillia Makedon. 2021. A survey of robots in healthcare. *Technologies* 9, 1 (2021), 8.
[18] Frank L Lewis, Draguna Vrabie, and Kyriakos G Vamvoudakis. 2012. Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. *IEEE Control Systems Magazine* 32, 6 (2012), 76–105.
[19] Jinying Lin, Zhen Ma, Randy Gomez, Keisuke Nakamura, Bo He, and Guangliang Li. 2020. A review on interactive reinforcement learning from human social feedback. *IEEE Access* 8 (2020), 120757–120765.
[20] Batta Mahesh. 2020. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]* 9, 1 (2020), 381–386.
[21] Margarita Mayo. 2017. The gender gap in feedback and self-perception. https://hbr.org/2016/08/the-gender-gap-in-feedback-and-self-perception
[22] José Del R Millán, Daniele Posenato, and Eric Dedieu. 2002. Continuous-action Q-learning. *Machine Learning* 49 (2002), 247–265.
[23] Andrew Y Ng, Stuart Russell, et al. 2000. Algorithms for inverse reinforcement learning.. In *Icml*, Vol. 1. 2.
[24] Elise Paradis, Bridget O'Brien, Laura Nimmon, Glen Bandiera, and Maria Athina Martimianakis. 2016. Design: Selection of data collection methods. *Journal of graduate medical education* 8, 2 (2016), 263–264.
[25] Martin L Puterman. 1990. Markov decision processes. *Handbooks in operations research and management science* 2 (1990), 331–434.
[26] Harish Ravichandar, Athanasios S Polydoros, Sonia Chernova, and Aude Billard. 2020. Recent advances in robot learning from demonstration. *Annual review of control, robotics, and autonomous systems* 3 (2020), 297–330.
[27] Reza Rawassizadeh, Taylan Sen, Sunny Jung Kim, Christian Meurisch, Hamidreza Keshavarz, Max Mühlhäuser, and Michael Pazzani. 2019. Manifestation of virtual assistants and robots into daily life: Vision and challenges. *CCF Transactions on Pervasive Computing and Interaction* 1 (2019), 163–174.
[28] Denise Rey and Markus Neuhäuser. 2011. *Wilcoxon-Signed-Rank Test*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1658–1659. https://doi.org/10.1007/978-3-642-04898-2_616
[29] Laurel D Riek. 2017. Healthcare robotics. *Commun. ACM* 60, 11 (2017), 68–78.
[30] Tomi-Ann Roberts and Susan Nolen-Hoeksema. 1989. Sex differences in reactions to evaluative feedback. *Sex Roles* 21 (1989), 725–747.
[31] Michael Schneier, Michael Schneier, and Roger Bostelman. 2015. *Literature review of mobile robots for manufacturing*. US Department of Commerce, National Institute of Standards and Technology ....
[32] Thomas B Sheridan. 2016. Human–robot interaction: status and challenges. *Human factors* 58, 4 (2016), 525–532.
[33] Sergio Spano, Gian Carlo Cardarilli, Luca Di Nunzio, Rocco Fazzolari, Daniele Giardino, Marco Matta, Alberto Nannarelli, and Marco Re. 2019. An efficient hardware implementation of reinforcement learning: The q-learning algorithm. *Ieee Access* 7 (2019), 186340–186351.
[34] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
[35] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning* 8 (1992), 279–292.
[36] Patrick Henry Winston. 1984. *Artificial intelligence*. Addison-Wesley Longman Publishing Co., Inc.
[37] Hang Yu and Elaine Schaertl Short. 2021. Active feedback learning with rich feedback. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 430–433.

9 APPENDIX

9.1 Questionnaire

Robot learning from rich feedback - Questionnaire

Robot Teaching

In this section, the focus is on gathering participants' feedback regarding their interaction with the robotic arm and their experience with the teaching process used in the experiment.

1. **Participant Number**
   (Given by experimenter)

   ____

2. **I found the teaching process intuitive to me and it was easy to follow.**

   Strongly Disagree ○—○—○—○—○ Strongly Agree

3. **In which situations did you decide to provide Demonstrations?**:

   ____

4. **I found it intuitive to provide Demonstrations.**

   Strongly Disagree ○—○—○—○—○ Strongly Agree

5. **In which situations did you decide to provide Evaluative Feedback?**:

   ____

6. **I found it intuitive to provide Evaluative Feedback.**

   Strongly Disagree ○—○—○—○—○ Strongly Agree

7. **How did you choose the starting positions of the agent (robotic arm)?**:

   ____

8. **The robotic arm learned effectively from my teaching inputs.**

   Strongly Disagree ○—○—○—○—○ Strongly Agree

9. **On a scale from 1-10 how would you rate your performance as a teacher?**

   Poor ○—○—○—○—○—○—○—○—○—○ Excellent

10. **What are the three words that come to mind when describing your overall experience with the interaction with the robotic arm?**:

    ____________ ____________ ____________

Using the robot everyday

The second section explores participants' willingness to integrate the robotic arm into their everyday tasks and their preferences for communication with the robotic arm in case they would be willing to teach it.

11. **Imagine you own a robotic arm, would you prefer to either personally teach it to perform and assist with everyday tasks, customizing its actions based on your preferences and needs, or have it pre-trained to assist with specific everyday tasks, ready to use right away without requiring personal teaching?**
    - ○ Personally teach it
    - ○ Pre-trained
12. **I would teach a robotic arm to help with:**
    - ☐ Cooking and meal preparation
    - ☐ Cleaning and housekeeping
    - ☐ Assisting with personal care (e.g., dressing, grooming)
    - ☐ Handling household chores (e.g., laundry, dishes)
    - ☐ Assisting with mobility or physical tasks
    - ☐ Other: ____________________
13. **To provide Demonstrations to a robotic arm, which way would you prefer to communicate with it?**
    - ○ Voice commands
    - ○ Touchscreen interface
    - ○ Physical gestures (e.g., hand movements - moving the robot)
    - ○ Smartphone app
    - ○ Other: ____________________
14. **To provide Evaluative Feedback to a robotic arm, which way would you prefer to communicate with it?**
    - ○ Voice commands
    - ○ Touchscreen interface
    - ○ Physical gestures (e.g., hand movements - moving the robot)
    - ○ Smartphone app
    - ○ Other: ____________________

Background Information

This section of this questionnaire aims to collect participants' background information. It includes several questions that provide insights into the participants' educational level, teaching experience, programming background, and experience with video games.

15. **Age**

    ____
16. **Gender**
    - ○ Male
    - ○ Female
    - ○ Other: ____________________
17. **What is the highest degree or level of school you have completed? If currently enrolled, highest degree received.**
    - ○ High School Diploma or Equivalent
    - ○ Associate's Degree or Technical/Vocational Certificate
    - ○ Bachelor's Degree
    - ○ Master's Degree
    - ○ Doctorate/Ph.D.

18. Have you ever owned a pet?

- ○ Yes
- ○ No

19. To what extent does your profession involve teaching or training other people?

None at all ○—○—○—○—○ To a great extent

20. To what extent do you have experience with programming?

None at all ○—○—○—○—○ To a great extent

21. To what extent do you have experience playing video games?

None at all ○—○—○—○—○ To a great extent