

# FARPLS: A Feature-Augmented Robot Trajectory Preference Labeling System to Assist Human Labelers' Preference Elicitation

Hanfang Lyu

hanfang.lyu@connect.ust.hk

Hong Kong University of Science and  
Technology  
Hong Kong, China

Ujaan Das

udas@connect.ust.hk

Hong Kong University of Science and  
Technology  
Hong Kong, China

Yingchi Li

terenceli@hkflair.org

Hong Kong Industrial Artificial  
Intelligence and Robotics Centre  
(FLAIR)  
Hong Kong, China

Yuanchen Bai

ybai2@andrew.cmu.edu

Carnegie Mellon University  
Pittsburgh, United States

Chuhan Shi

cshiq@connect.ust.hk

Southeast University  
Nanjing, China

Mingfei Sun

mingfei.sun@manchester.ac.uk

Department of Computer Science,  
University of Manchester  
Manchester, United Kingdom

Xiaojuan Ma\*

mxj@cse.ust.hk

Hong Kong University of Science and  
Technology  
Hong Kong, China

Xin Liang

XinLiang0920@gmail.com

Tongji University  
Shanghai, China

Leiliang Gong

leiliang@hkpc.org

Robotics and AI Division, Hong Kong  
Productivity Council  
Hong Kong, China

Ming Ge

mingge@hkpc.org; mingge@hkflair.org

Hong Kong Productivity Council;  
Hong Kong Industrial Artificial  
Intelligence and Robotics Centre  
(FLAIR)  
Hong Kong, China

## ABSTRACT

Preference-based learning aims to align robot task objectives with human values. One of the most common methods to infer human preferences is by pairwise comparisons of robot task trajectories. Traditional comparison-based preference labeling systems seldom support labelers to digest and identify critical differences between complex trajectories recorded in videos. Our formative study ( $N = 12$ ) suggests that individuals may overlook non-salient task features and establish biased preference criteria during their preference elicitation process because of partial observations. In addition, they may experience mental fatigue when given many pairs to compare, causing their label quality to deteriorate. To mitigate these issues, we propose FARPLS, a Feature-Augmented Robot trajectory Preference Labeling System. FARPLS highlights potential outliers in a wide variety of task features that matter to humans and extracts the corresponding video keyframes for easy review and comparison. It also dynamically adjusts the labeling order according to

users' familiarities, difficulties of the trajectory pair, and level of disagreements. At the same time, the system monitors labelers' consistency and provides feedback on labeling progress to keep labelers engaged. A between-subjects study ( $N = 42$ , 105 pairs of robot pick-and-place trajectories per person) shows that FARPLS can help users establish preference criteria more easily and notice more relevant details in the presented trajectories than the conventional interface. FARPLS also improves labeling consistency and engagement, mitigating challenges in preference elicitation without raising cognitive loads significantly.

## CCS CONCEPTS

- Human-centered computing → HCI design and evaluation methods; Interactive systems and tools.

## KEYWORDS

data labeling, preference collection system, human-robot value alignment.

## ACM Reference Format:

Hanfang Lyu, Yuanchen Bai, Xin Liang, Ujaan Das, Chuhan Shi, Leiliang Gong, Yingchi Li, Mingfei Sun, Ming Ge, and Xiaojuan Ma. 2024. FARPLS: A Feature-Augmented Robot Trajectory Preference Labeling System to Assist Human Labelers' Preference Elicitation. In *29th International Conference on Intelligent User Interfaces (IUI '24), March 18–21, 2024, Greenville, SC, USA*, <https://doi.org/10.1145/3640543.3645145>

\*Corresponding author

## 1 INTRODUCTION

Advancement in artificial intelligence (AI) and robotics technologies brings robots out of laboratories, requiring them to perform daily tasks for or with humans [47, 54, 56, 58, 67]. To accommodate this requirement, robot task learning aims at teaching robot tasks such as prehabilitation [81], companion [64], assembly tasks [61] according to user preferences. Traditional robot learning algorithms rely on delicately handcrafted reward functions to guide robot behavior. However, these delicate reward functions may not accurately reflect humans' true values [18] due to generalization errors [57], task misspecifications [14], etc. The human-robot value alignment can not only improve robot performances according to humans' preference but also avoid undesired robot behavior and even safety issues [6, 68, 83]. Learning a reward model from human preferences hence emerges [18], which leverages the computationally efficient and user-friendly pairwise comparison to collect human preferences [8, 14, 38, 39]. However, this learning process still requires a substantial number of high-quality human preference inputs, which inevitably contain subjective uncertainties and incur a huge cognitive labor cost for participants. Reducing the cost and uncertainty in the human preference data collection process has been one of the focuses of research on developing robots for human use.

The human preference collection process includes recruiting adequate human labelers to give high-quality preference annotations to many robot task trajectories. The annotation requires human labelers to understand every robot trajectory presented and to specify which trajectory is better via comparison from their point of view. Previous studies mainly handled the challenges in human label collection from two algorithmic directions. One line of research proposed to work with human data with inconsistent qualities, trying to incorporate human uncertainty into active reward learning [38]. The other line of work focused primarily on reducing the number of human labels needed, for example, by prompting pairs with a high information gain [5]. These studies, however, largely overlooked possibilities of improving data quality and human engagement by assisting in labelers' sensemaking process during preference elicitation. Due to the complexity of full episodes of robot trajectory data, understanding and comparing nuanced characteristics of robot task processes and performance based on trajectory-recording videos may be difficult for labelers, especially those without robotics expertise. Consequently, human labelers may find the trajectories presented to them arbitrary or misleading [35], leading to inconsistent labeling and low-quality preference data. Thus, supporting human labelers' efficient labeling from novel interaction and design is an important aspect besides algorithmic perspectives.

In this work, we argue that the human preference collection system for robot task learning should not only focus on the algorithms' perspective of reducing the cost and uncertainties but also provide sufficient support to human labelers and improve the data quality from the source. We aim to address the following three research questions:

**RQ1** How do human labelers compare robot task trajectories, and what are their challenges and needs when eliciting their trajectory preferences?

**RQ2** How to design a preference collection system that can assist human labelers in trajectory sensemaking and preference elicitation?

**RQ3** How does the proposed system improve data quality and human engagement and mitigate the challenges above?

In this paper, we gain an understanding of how human labelers usually compare two robot task trajectories through semi-structured interviews in a formative study with 12 participants. We derive a list of trajectory features humans would consider as preference elicitation criteria by combining findings from previous literature and the formative study. We also identify three main challenges that human labelers face in the preference collection process: difficulty in forming criteria, overlooking trajectory details, and difficulty in maintaining focus, which align with data labeling challenges in other fields (e.g., [30, 41]).

Drawing on the human needs and derived design requirements, we designed FARPLS, a Feature-Augmented Robot trajectory Preference Labeling System to address the challenges. We generate a dataset of robot trajectories of a typical pick-and-place task commonly seen in human environments and automatically extracted the features relevant to human criteria from the trajectories. We clustered the trajectories based on these features and designed a prompting strategy to initially present the trajectory pairs that are varied in features to facilitate criteria formation. FARPLS is a web-based system that allows human labelers to compare two robot trajectories pairwise. The system marks feature-based keyframes in the video for accessible replay and comparison and highlights features that defer the most from the mean. The system also provides real-time attention monitoring and feedback to help human labelers maintain attention and engagement.

We conducted a between-subjects study with 42 participants to compare the proposed tool with the conventional pairwise comparison interface. Each participant was required to label 105 pairs of robot pick-and-place trajectories and fill out a post-study questionnaire regarding confidence, cognitive load, and challenges from the formative study. The results show that FARPLS significantly improves the labeling consistency of human labelers without raising their cognitive loads. The subjective ratings from the questionnaire also show that FARPLS is significantly better than the baseline in terms of establishing comparison criteria and providing sensemaking supports to help labelers notice more nuanced details. Although the average labeling time of labelers using FARPLS is significantly longer than those with the baseline system, the participants' perceived engagement in rewards and boredom in the labeling process is significantly improved.

The key contributions of our work are as follows:

- We conducted a formative study to understand human labelers' pairwise preference elicitation process and their needs and challenges.
- We propose a novel trajectory preference labeling system for robot manipulation tasks, FARPLS, that can help human labelers give high-quality preference data.
- We conducted a user study to evaluate the effectiveness of our system and provide design considerations for future data collection and reward learning systems.

## 2 RELATED WORK

### 2.1 Human Preference Learning in Robot Manipulation Tasks

Recent research has pointed out the potential misalignment issue between human values and robotic objectives. Studies to address this issue include bidirectional human-robot communication in group settings [83], evaluation of task accomplishment [6], and disentangled representation learning (DRL) [77]. Particularly, Reinforcement Learning from Human Preference (RLHP) [1, 14, 49, 50, 84] emerges as a new trend to offer a flexible and adaptable way to fine-tune an agent's behavior based on human preference. RLHP comes with three key steps: (1) human preference collection, (2) reward learning, and (3) RL policy optimization. Human preference has been collected in various ways including absolute rating [13], and ranking (voting [20], pairwise comparison [33, 46, 63], multiple ranking [9, 10, 60, 89], etc.). Among them, pairwise comparison is the most widely used due to its advantages in terms of its impacts on model performance [40, 43] and labeler experience [8, 38].

These existing studies on human preference collection and learning primarily focus on the trajectory querying strategy, human data augmentation and representation [14], with little attention paid to assisting the human preference labeling process, especially on how human values can be aligned in robotic trajectory ([4, 59]) for better labeling outcomes. We focus on the robotic trajectory evaluation by exploring key features that are deemed valuable in the robotics community (e.g., [25, 51, 74, 82, 88]). We present a system that conveys the feature information in a comprehensible manner to labelers with different levels of prior knowledge, especially those non-expert labelers. Further, our study identifies the multifaceted challenges of robotics trajectory preference labeling, thus offering a chance to optimize the entire crucial procedure and its subsequent outcomes.

### 2.2 Data Annotation Tools for Machine Learning

Labeling tools have been designed to deal with various tasks and data types, such as text (AILA [17], DUALIST [69]), image (EasyAlbum [22], SAPHARI [72]), video (MediaTable [65], VoTT [55]), audios (VIA [28]), and other special use cases like activity label from the elderly [42]. Three main branches of various techniques include semi-automatic labeling, active learning, and novel interaction and visual design [87]. Active learning (e.g., ALVA [45], DUALIST [69], MI3 [86]) and semi-automatic labeling (e.g., ISSE[11], V-aware[32]) are widely applied to optimize the labeling process and outcomes from the algorithmic perspective. Novel designs of the interfaces are essential in supporting efficient labeling, such as a combination of tabular and bucket list [65], keywords highlighting [17], and instances clustering [22].

Existing labeling tools are all designed for specific data types, and they fail to tackle the unique challenges of robotic trajectory labeling since robotic arm trajectory introduces unique complexities compared to the data types mentioned above. The trajectory data are highly non-linear and non-stationary, and the human labelers may not have a clear understanding of their movements and effects on the environment. We focus on understanding the trajectory

labeling process and designing a labeling system to facilitate human labeling of robotic arm trajectories.

### 2.3 Visualization in Human-Robot Interaction

In the realm of human-robot collaboration, significant efforts have been directed toward enhancing the visualization and demonstration of robot tasks, as evidenced by previous research studies (e.g., [2, 16, 21]). To provide some more detailed examples: Chandan et al. [15] develops an intelligent augmented reality (AR) agent that learns visualization policies aimed at enhancing efficiency and minimizing distractions for humans; Dragan et al. [26] enhances the interface for human communication with a virtual robot and improves the robot's knowledge representation through the use of 3D isometric visualization, along with providing the robot's first-person perspective; Zhu and Veloso [90] targets the challenge of visual mismatch with video capture, aiming to facilitate the overlay of visualizations onto video streams and extract the underlying algorithms utilized; Szafir and Szafir [73] proposed a data-centric HRI framework and identified visualization design concepts to facilitate HRI data tasks such as data collection, analysis, and human decision-making.

Existing interfaces for HRI systems primarily focus on situation awareness and user control [73, 76]. In these contexts, the interfaces focus on how to assist the human-robot interaction and data analysis process and fail to consider how to improve the human experience in the labeling process and the data quality. It remains unclear what specific information is required and how it should be presented within the context of providing preferences for the robot arm manipulation tasks. In our work, we will draw inspiration from previous interface design and data presentation approaches, incorporating novel insights from a formative study to enhance the interaction between human labelers and our labeling system for robot arm trajectories.

## 3 FORMATIVE STUDY

In this section, we present the formative study and findings to answer RQ1. With the Institutional Review Board (IRB) approval from the University Research Ethics Committee, we conducted a formative study to explore how labelers compare robot trajectories and figure out challenges and needs in traditional pairwise trajectory preference collection systems.

### 3.1 Study Design and Procedure

**3.1.1 Participants.** We recruited twelve participants (6 females and 6 males) aged 19 to 28 ( $M = 22.0$ ,  $SD = 2.6$ ) to participate in the formative study. Their familiarity with the data labeling system (1 for Not familiar at all – 5 for Extremely familiar) ranges from 1 to 3 ( $M = 1.9$ ,  $SD = 0.9$ ). Additionally, their familiarity with robotics (1 for Not familiar at all – 5 for Extremely familiar) ranges from 1 to 4 ( $M = 1.9$ ,  $SD = 0.9$ ).

**3.1.2 Materials.** We built a basic preference labeling system following the interface of Christiano et al. [18] shown on their project website <sup>1</sup>. We generated a pick-and-place dataset based on the *Can*

<sup>1</sup><https://openai.com/research/learning-from-human-preferences>

task of robomimic<sup>2</sup>, with six cans on the table and four bins on the side. As shown in Figure 1, six cans are randomized on the left table, and the robot’s goal is to pick up one of the cans and place it into the bins on the right. A robot trajectory consists of states and actions, where the state is a vector of observations at each step (i.e., each frame in the video), including the positions and velocities of all joints of the robot arm and all the objects collected using the robosuite framework [91]. The states are recorded in each frame, and the trajectory videos are recorded at 20 frames per second. In the basic preference labeling system, we utilized the K-means [37] to perform cluster analysis on the states. This analytical approach yielded 9 representative samples from each cluster, and the system randomly prompts the 36 trajectory pairs during the study session.

**3.1.3 Procedure.** After getting familiar with the pick-and-place task and the labeling system, the participants labeled 36 pairs of robot pick-and-place trajectories through the system with their screen recorded. The total labeling time ranges from 6 to 22 minutes ( $M = 15.4$ ,  $SD = 4.0$ ). After the labeling session, we conducted a retrospective think-aloud study [29] to let the participants recall their mental process using the screen recording, so that we can understand the way that they compare the trajectories. Finally, we conducted semi-structured interviews with the participants to gain insights into: (a) the challenges they encounter when using the trajectory comparison system, (b) their criteria and features for comparing robot trajectories, (c) potential additional challenges they may face in scenarios involving labeling more pairs, and (d) their requirements for the trajectory preference labeling system. The linked document<sup>3</sup> presents the details of the formative study including demographic questions, procedures, and semi-structured interview questions.

All the study sessions were fully recorded and transcribed for thematic analysis. Two researchers independently coded the transcripts and got 22 initial codes under 3 themes, and we discussed the coding results in our research team meetings to categorize or divide the initial codes into 9 codes and 19 subcodes and reach a consensus. Three themes were identified from the thematic analysis: (1) what are the criteria and corresponding features that labelers care about in comparing robot trajectories, (2) challenges in the preference elicitation process, and (3) needs of the trajectory preference labeling system. The formative study’s findings are summarized in the following subsections.

## 3.2 Criteria and Features

Participants compared two robot task trajectories based on observed criteria and features in the trajectory videos, which sheds light on the first part of RQ1: how labelers compare robot trajectories. While prioritizing these criteria and features is challenging (described in Section 3.3.1), most participants considered the criteria and corresponding features in a hierarchical order. During our interviews, we asked participants to suggest additional criteria and features they would consider for comparing trajectories other than those mentioned in the retrospective think-aloud. After that, we presented a list of criteria and features summarized from other literature, such as related human values in [83] and evaluation metrics from [25].

<sup>2</sup>[https://robomimic.github.io/docs/v0.2/datasets/robomimic\\_v0.1.html#info](https://robomimic.github.io/docs/v0.2/datasets/robomimic_v0.1.html#info)

<sup>3</sup><https://bit.ly/3Hzil21>

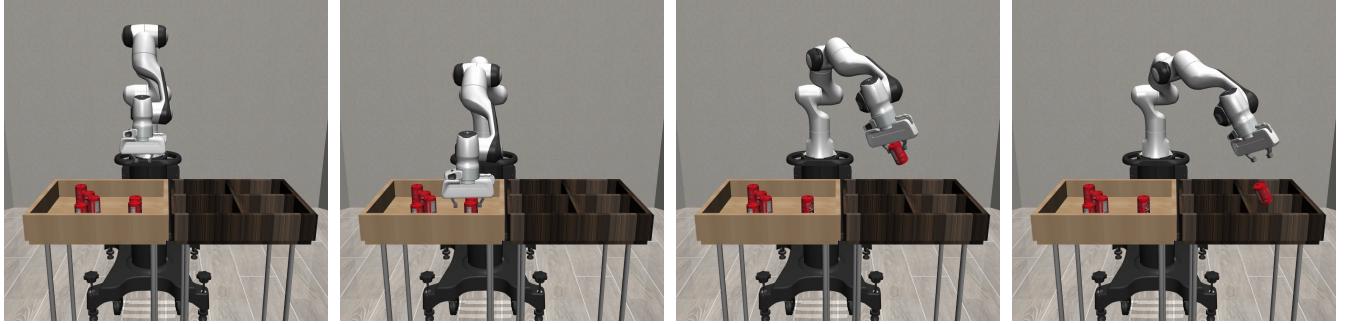
Some participants (P01, P03, P05, P12) admitted they would not consider certain specialized robot arm metrics from [25], finding them hard to interpret or not important from their perspective. For example, P01 ignored the arm and the tilting angle and cared more about whether the object was safe or the task was completed within an acceptable time, etc.. Similarly, P12 expressed his perspective, stating that judging the system’s industrial performance might not be straightforward and could be subjective based on their intuitive point of view. The way that participants compare the trajectories can reflect the mismatch between the preferences of common users (from high-level criteria) and the evaluation metrics in the robotic specializations (from low-level metrics). Thus, rather than using features obtained directly from metrics in the previous literature, we summarized a set of criteria and corresponding features mentioned by labelers in our thematic analysis in Table 1.

These features are summarized using generalizable terms that can be applied to other robot tasks. For example, in this pick-and-place task, the *distance* feature can be measured by the highest distance between the can and the table and the nearest distance between the can and the table edge, etc.. All participants mentioned either *safety* or *efficiency* as their top priority criteria. The *safety* criterion is related to *collision*, *distance*, and *contact force* features. The *efficiency* criterion is related to *speed*, *path length*, *time*, and *power usage* features. For *power usage* and *contact force*, they are not observable from the trajectory videos, but participants mentioned that they would consider them if provided with the information. The participants recognize the task quality criterion less, and it is related to *speed smoothness*, *trajectory smoothness*, *orientation*, and *grasp position* features. However, participants talked about several features related to the task quality criterion in their think-aloud, such as smoothness (P01, P03, P04, P05, P08, P12), stableness (P06, P07, P08), object orientation (P02, P05, P06, P11, P12), gripper orientation (P06, P09), or grasp approach (P02, P03, P08, P11, P12). For example, P06 thought *lying flat down would be better* when comparing the can orientation of two trajectories.

## 3.3 Challenges

This subsection answers RQ1 about the challenges during the preference elicitation process. We categorize participants’ challenges when comparing robot trajectories into three aspects (Figure 2’s “Challenges” column).

**3.3.1 C1 - Difficulty in forming criteria.** Forming criteria for comparing robot trajectories is the first step of the preference elicitation process, which is challenging for participants from the following two aspects: *C1.1 - unclear coverage of criteria features* and *C1.2 - unclear value distribution of each feature*. On the one hand, participants (P01-P07, P12) knew too little about the performance of this robot to build a set of criteria covering all the new situations. For example, P05 first felt that *one of the essential things was that the robot could not knock the can down, and the robot probably will not be able to grab it next time*. But then he found that the robot could push the can down and grab it, which made him change the priority of the feature *collision*. On the other hand, participants (P02, P05, P08, P09) found it difficult to estimate the value distribution of each feature. Thus, they were not sure about the importance of each feature. For example, many participants found it difficult to



**Figure 1:** The pick-and-place task for preference labeling from the *frontview*. The robot picks up a can from the left table and places it on the target table with four bins.

**Table 1:** Criteria and corresponding features that labelers consider in comparing robot trajectories

Criterion	Feature	Description
Safety	Collision	Contact(s) of cared objects.
	Distance	The distance(s) between the cared point(s) / line(s) / surface(s).
	Contact force	The contact force(s) between the cared objects.
Efficiency	Speed	The speed(s) of the cared object(s).
	Path Length	The length(s) of the path(s) of the cared object(s).
	Time	The total time and time used for cared path(s).
Task Quality	Power Usage	The total power consumption.
	Speed Smoothness	The sum(s) of absolute changes in the speed(s) of the cared object(s).
	Trajectory Smoothness	The smoothness score(s) of the path(s) of the cared object(s).
Orientation	Orientation	The relative orientation(s) between the gripper(s) and the target(s).
	Grasp Position	The relative position(s) between the gripper(s) and the grasped target(s).

estimate the value distribution of the feature *path length* because they did not know the range of the path length of the gripper or the can, which affected the priorities of the criterion *efficiency*.

**3.3.2 C2 - Overlooking trajectory details.** Participants found it challenging to pay attention to the details of the trajectories. The challenge includes two factors: *C2.1 - the lack of robotic understandings* and *C2.2 - the unobservability of video details*. Due to the knowledge limitation, participants (P01, P02, P03, P06, P09, P12) found figuring out some robotic details challenging. For example, P03 admitted not having enough knowledge about the robotic arm to judge its power usage without knowing which joint is more power-consuming. Besides, observing the non-salient details of the trajectories from the videos is also not accessible and some features like *contact force* are not observable from the trajectory videos (Mentioned by P01-P03, P05-P07). For example, during retrospective think-aloud, P07 did not notice a minor collision between the can and the table.

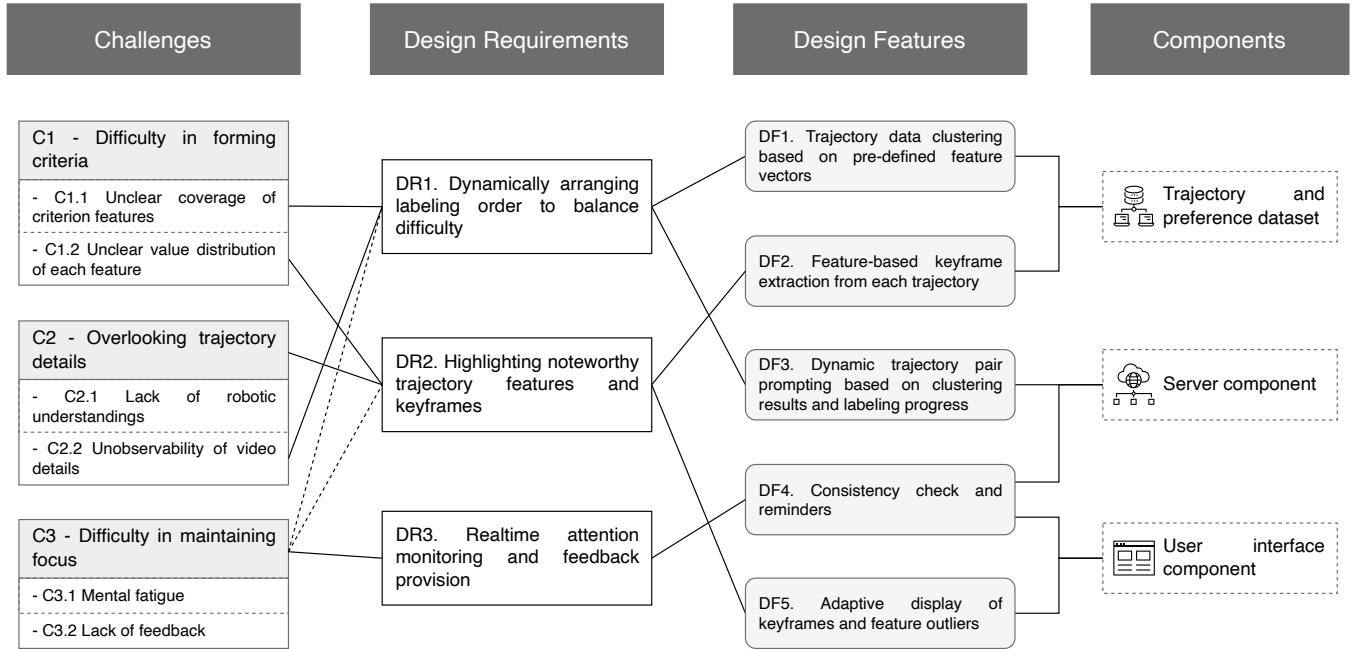
**3.3.3 C3 - Difficulty in maintaining focus.** The labeling process is time-consuming, and many participants failed to maintain focus during the labeling process. The following two factors contribute to this challenge: *C3.1 - mental fatigue* and *C3.2 - lack of feedback*.

Labeling similar trajectories showing the same simple task, participants (P01, P02, P04) felt bored and mentally exhausted. For example, P04 said that “*I’m going to be honest with you, this was incredibly tedious.*” Additionally, participants (P03, P05, P07) anticipated knowing their progress, the quality of their labeling, and improvements in the robot’s performance but did not receive any feedback, which hindered their ability to stay focused. Many participants kept asking how many pairs were left during the labeling process.

### 3.4 Derived Design Requirements

The last part of RQ1 is what labelers need for the trajectory preference labeling system. We derive the design requirements and features of the trajectory preference labeling system from the coded challenges and needs. The design requirements and features are in the “Design Requirements” and “Design Features” columns of Figure 2, with the relations between columns shown in the connecting lines.

**3.4.1 DR1. Dynamically arranging the labeling order to balance difficulty.** DR1 aims to address the challenges of C1.1 and C2.2, which



**Figure 2: This figure demonstrates an overview of our design pipeline: from identified challenges, we establish the design requirements to address each challenge, then design corresponding features to meet those requirements, and finally, we integrate those features into concrete components of our labeling system.**

involve helping labelers establish a consistent criteria system and manage the difficulty of labeling. Participants found it challenging to label similar trajectory pairs without establishing consistent criteria. Therefore, they wanted to label trajectory pairs with distinct features to establish consistency at the beginning of the labeling task. To address this challenge, DR1 requires the system to prompt labelers to label trajectory pairs with distinct features at the initial stage, followed by trajectory pairs with similar features after labelers have viewed a more comprehensive range of feature varieties. Furthermore, we propose design features *DF1. Trajectory data clustering based on pre-defined feature vectors* and *DF3. Dynamic trajectory pair prompting based on clustering results and labeling progress* to satisfy DR1.

**3.4.2 DR2. Highlighting noteworthy trajectory features and keyframes.** The goal of DR2 is to address the challenges of C1.2 and C2, i.e., help labelers identify features and keyframes of trajectories. More information about features not easily identified from the videos, such as *power usage* and *contact force*, can help participants prioritize the features more accurately. Participants also wished for the ability to jump to specific keyframes to aid in recalling trajectory features and observing details. To fulfill this need, DR2 requires the system to provide more information, including distributions of non-salient features and keyframes, to assist labelers in making comparisons. We propose design features *DF2. Feature-based keyframe extraction from each trajectory* and *DF5. Adaptive display of keyframes and feature outlier distributions* to satisfy DR2.

**3.4.3 DR3. Real-time attention monitoring and feedback provision.** DR3 is to tackle the challenge C3, which is to help labelers focus

on the labeling task and provide feedback to labelers. Participants mentioned that they would like to know how much progress they have made and how well they are doing. Thus, we use participants' consistency to proxy their attention and fatigue to monitor their performance and provide feedback. To realize DR3, *DF4. Consistency check and reminders* is proposed to monitor labelers' attention and provide feedback to participants.

## 4 FARPLS: SYSTEM DESIGN AND IMPLEMENTATION

This section provides a detailed illustration of the design and implementation of our FARPLS system, which comprises three main subsections: the dataset, the server component, and the user interface component (Figure 2's "Components" columns). These components are implemented to fulfill design features as shown in lines connected to "Design Features" in Figure 2.

### 4.1 Feature-Augmented Trajectory Dataset

**4.1.1 Dataset generation.** The pick-and-place task is common in real life and thus a good starting point for generating a dataset for preference labeling. We create the simulation environment using the robosuite framework [91]. The *Can* task of robomimic [52] dataset<sup>4</sup> contains 200 successful proficient human demonstration trajectories (ph), 300 successful multi-human demonstration trajectories (mh), 3900 machine-generated trajectories (mg) and 100 successful human demonstration trajectories paired with 100 unsuccessful human demonstration trajectories (paired). As the

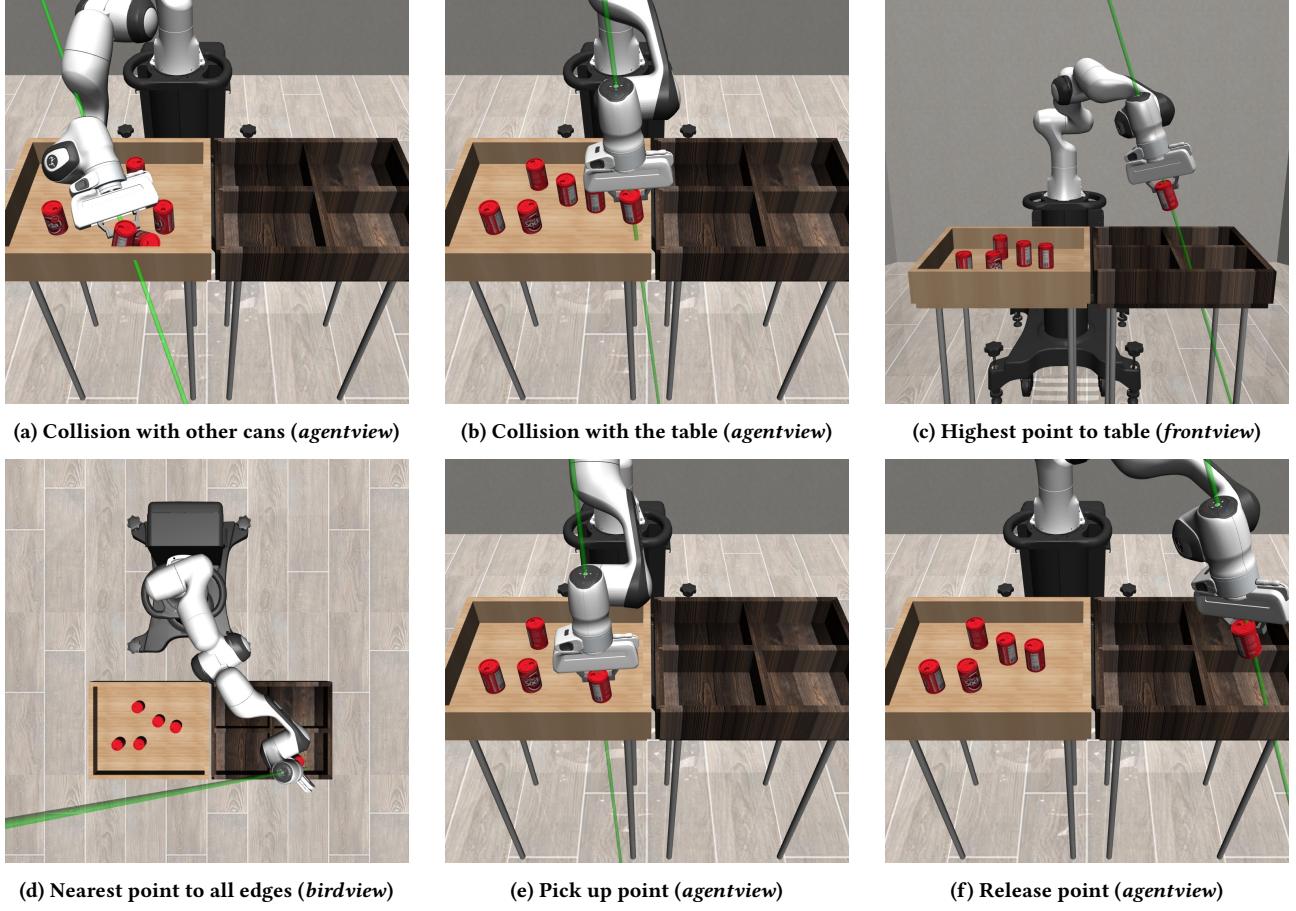
<sup>4</sup>[https://robomimic.github.io/docs/v0.2/datasets/robomimic\\_v0.1.html#info](https://robomimic.github.io/docs/v0.2/datasets/robomimic_v0.1.html#info)

**Table 2: Definitions for each feature and corresponding keyframes.**

Feature	Definition	Keyframe
Safety		
Collision	The number of contacts between all the objects, except those between robot fingers and the target can.	The frames with the collisions (Figures 3a and 3b)
Distance	1) The minimum distances between the target can and each table edge. 2) The maximum height of the target can from the table surface.	The frames of the can: 1) with the minimum distance to all table edges (Figure 3d); 2) with the maximum height to the table (Figure 3c)
Contact force	The maximum force exerted by the robot end-effector.	N.A.
Efficiency		
Speed	The average speed of the robot end-effector during the task.	N.A.
Path Length	1) The path length of the robot end-effector reaching the target can. 2) The path length of the end-effector grasping the target can. 3) The path length of the can from pick-up to placement.	N.A.
Time	1) The time of the robot end-effector reaching the target can. 2) The time of the end-effector grasping the target can. 3) The time from pick-up to placement. 4) The total time of the manipulation task.	The frames of: 1) the pick-up point (Figure 3e); 2) the release point (Figure 3f).
Power Usage	The sum of the absolute values of the joint rotations as a proxy for power usage [25], i.e., $pseudo\_cost = \sum_{i=1}^n  q_i ,$ where $q_i$ is the joint rotation of the $i$ -th joint, and $n$ is the number of joints.	N.A.
Task Quality		
Speed Smoothness	The sum of the absolute values of the end-effector's acceleration, i.e., $speed\_smoothness = \sum_{j=1}^s \sqrt{a_j^2},$ where $a_j$ is the 6-dimension acceleration vector of the end-effector at the $j$ -th state and $s$ is the number of states in the trajectory.	N.A.
Trajectory Smoothness	The sum of angles between the displacements between adjacent states, i.e., $trajectory\_smoothness = \sum_{j=1}^s \arccos \frac{\mathbf{x}_j \cdot \mathbf{x}_{j+1}}{ \mathbf{x}_j  \mathbf{x}_{j+1} },$ where $\mathbf{x}_j$ is the 3-dimension end-effector displacement vector between the $(j-1)$ -th state and the $j$ -th state.	N.A.
Orientation	The maximum relative angle between the can's and the end-effector's orientation.	N.A.
Grasp Position	The relative position vector with the largest distance between the can's and the end-effector's center during the grasping time.	N.A.

robomimic *Can* task contains only a single can, we randomly simulated the initial positions of the other five cans on the left table as a new environment. First, we use the actions from the robomimic dataset to generate successful trajectories for the robot to pick up and place one of the cans in our environment. Then, we select trajectories where the robot's end-effector remained confined within the horizontal spatial constraints of the edges of two tables. Furthermore, we remove human demonstration trajectories longer than 8

seconds, which is the longest duration in machine-generated trajectories. Finally, we get our dataset *PickPlaceCans* with 636 successful trajectories, including 200 ph, 117 mh, 99 paired and 220 mg. All participants in the formative study expressed a preference for the *frontview*, i.e., the camera view from the front, over other views. Therefore, we continued to use the *frontview* videos to demonstrate the trajectories.



**Figure 3: Examples of five kinds of keyframes, each captured from the best view to observe the corresponding feature.**

**4.1.2 Feature and keyframe definitions.** According to the general criteria and features that labelers care about in Section 3.2, we define each feature and feature-based keyframes to be extracted for each trajectory in *PickPlaceCans* in Table 2. The detailed formulas for each feature in either time series or scalar values are in Table 9 of Appendix A. We stack the time series features together to form the criterion vector series, so that we can represent each criterion for the trajectory and calculate the similarity between trajectories for clustering. We calculate the scalar values and extract keyframes to display the feature distribution on the interface, which is more intuitive for the labelers to compare the features of different trajectories.

**4.1.3 Criteria-based clustering.** We create a criterion vector series for each criterion with the time series of corresponding features stacked together.  $safety_i(t)$ ,  $efficiency_i(t)$ , and  $task\_quality_i(t)$  denote the criterion vector series of three criteria, *Safety*, *Efficiency*, and *Task Quality*, for trajectory  $i$  at  $t$ -th step (Appendix A). We compute Dynamic Time Warping (DTW) distance matrices [66]  $D_{safety}$ ,  $D_{efficiency}$ , and  $D_{task\_quality}$ . Each entry  $D_{safety}(i, j)$  denotes the DTW distance between the safety vector series of trajectory  $i$  and  $j$ .  $D_{efficiency}(i, j)$  and  $D_{task\_quality}(i, j)$  are defined similarly. Then we obtain the pairwise distance matrix  $D$  using the

weighted sum of the DTW distance matrices for each criterion to represent the similarity between trajectories:

$$D = \frac{w_s D_{safety} + w_e D_{efficiency} + w_t D_{task\_quality}}{w_s + w_e + w_t},$$

We set equal weights for all criteria in this work. In practice, designers can add weights to prioritize the criteria in the similarity calculation. Then we obtain our feature-augmented, clustered dataset *PickPlaceCans*.

## 4.2 Server Component

**4.2.1 Prompting strategy.** We use the metrics in Table 3 to calculate a dynamic ranking score of all trajectory pairs for each labeler. These metrics are dynamically transformed to rank-based scores [80] to be averaged to the final ranking score. The larger the ranking score, the higher the priority of the trajectory pair to be prompted to the user.

At the initial stage, when the users have not seen all the clusters, the server chooses all trajectory pairs with none of their clusters covered and ranks them based on the mean of other metrics to prompt the user. This initial stage aims to help the user to have a whole glance at the features from different clusters, reducing

**Table 3: Metrics for dynamic ranking scores**

Metric	Definition	Ranking
<i>Cluster Coverage</i>	The portion of prompted trajectories from each cluster.	Descending
<i>Combination Familiarity</i>	The portion of labeled trajectory pairs from each cluster combination.	Descending
<i>Pair Similarity</i>	The distance between the normalized feature vectors (Section 4.1.2) from each trajectory pair.	Ascending
<i>Pair Disagreement</i>	The variance [18, 71] of all users' preference scores ( $preference\_score \in \{0, 0.5, 1\}$ ) for each trajectory pair.	Ascending
<i>Cluster Disagreement</i>	The variance of all users' preference scores ( $preference\_score \in \{0, 0.5, 1\}$ ) for all trajectory pairs in each cluster combination.	Ascending
<i>Label Skewness</i>	The number of users that labeled each trajectory pair.	Descending

the difficulty in forming criteria (*C1*). After the user has seen all the clusters, the server will choose the trajectory pairs with the least label skewness score and rank them based on other metrics to prompt the user. These metrics balance the labeling difficulty, user familiarity and disagreements (*DR1*).

**4.2.2 Consistency checking.** We arrange a consistency-checking round every 10 normal prompting rounds, except that the first consistency-checking round is arranged after prompting 15 unique trajectory pairs. For each consistency-checking round, we randomly select one labeled trajectory and prompt it to the labeler to check the consistency. If the labeler's preference label is consistent with the previous preference, we will prompt an encouraging message to the labeler:

According to our record so far, you have been rather careful and thorough in the past labeling sessions!  
Good job! Take a break if needed and keep on the good work.

Otherwise, we will prompt a message to remind the labeler to take a rest and be more careful:

Feeling tired? Take a break if necessary and please stay attentive in the following sessions.

This consistency-checking mechanism follows *DF4*. *Consistency check and reminders* aiming to fulfill *DR3*. *Real-time attention monitoring and feedback provision* that deals with challenge *C3 - Difficulty in maintaining focus*.

### 4.3 User Interface Component

We design the interface of FARPLS based on the interface of Christiano et al. [18] and add our design features related to the user interface component (*DF4* and *DF5*) The user interface overview and descriptions for all components are in Figure 4.

**4.3.1 Adaptive display of feature-based keyframes (*DF5*).** To facilitate the comparison of the feature-based keyframes extracted from each trajectory, we render keyframe buttons dynamically based on information sent alongside the keyframe to highlight information (upon hover), including the keyframe's start/stop time and an annotated thumbnail. For features wholly unique to each trajectory, such as "Collisions", the buttons are labeled as per the collision number

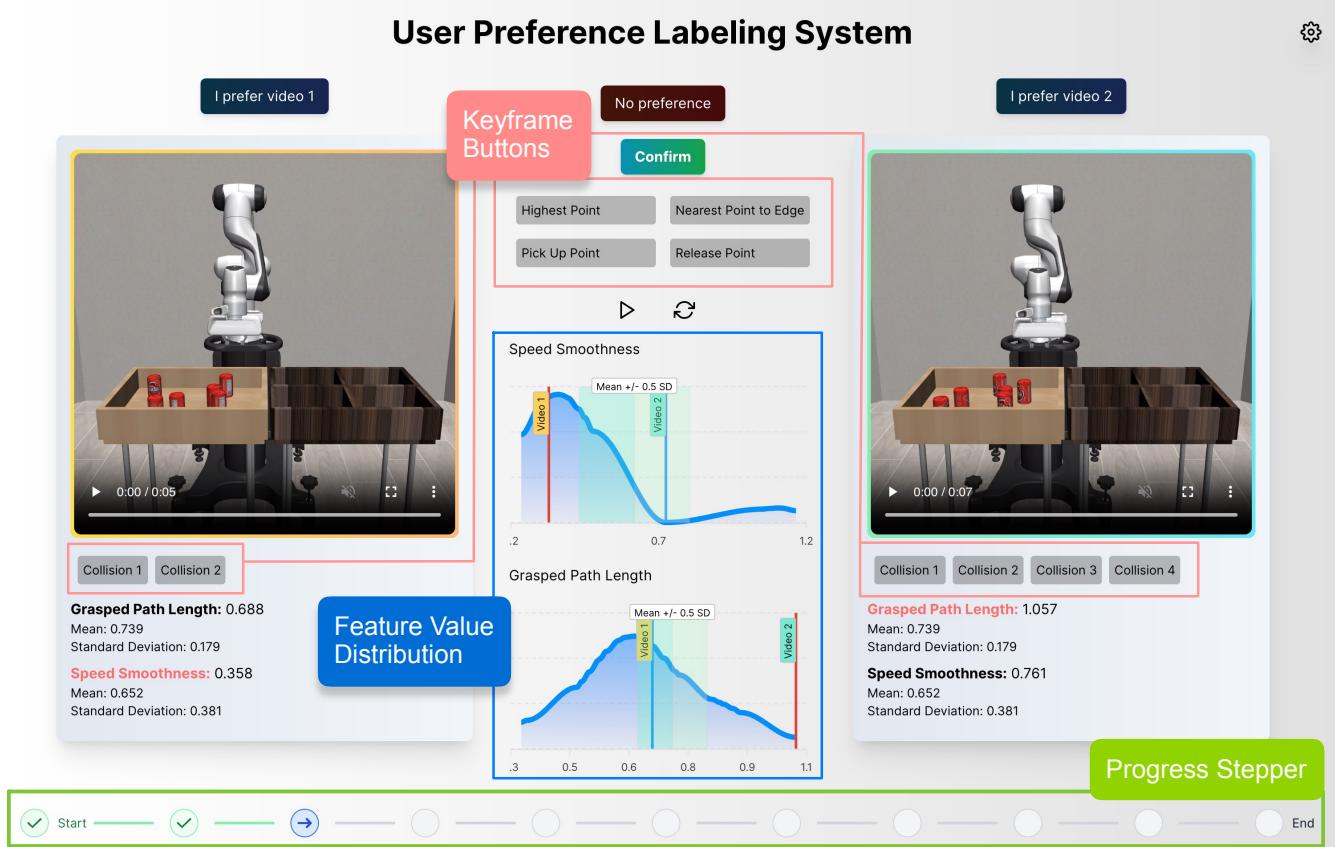
(i.e., "Collision 1", "Collision 2", "Collision 3", etc.), and placed underneath (Figure 5). For features shared between both videos, such as the "Pick Up Time" or "Highest Point" features, we label buttons appropriately, and on hover, show the keyframe information for both videos side-by-side (Figure 6). Upon clicking these buttons, the system loops the video (if it is a unique keyframe) or two videos (if it is a shared keyframe) between the start and stop time-frames to allow users to compare both features directly, thereby helping labelers to compare noteworthy differences between the two videos.

**4.3.2 Feature distribution visualization (*DF5*).** In addition, we visualize the feature distribution via density area charts to provide more context regarding the whole picture of the feature in the dataset. For each trajectory we render a density chart of the "outlying" feature, i.e., the feature with the maximum absolute value of its z-score [23], i.e.,  $|\frac{x-\mu}{\sigma}|$ . Additionally, the following values were illustrated and color-coded on the graph with easily identifiable colors for the overall statistics or to match the corresponding trajectories:

- the mean  $\mu$  with the range of plus and minus half std  $\sigma$ , i.e.,  $\mu \pm 0.5\sigma$ , in transparent green area to compare the given feature values to the average across the dataset,
- the outlying feature value in red line to highlight, and
- the same feature but from the other trajectory of the pair in blue line for comparison.

The interface typically displays two density charts, each highlighting an outlying feature from one trajectory in the pair. If both trajectories in the pair share the same outlying feature, the interface will only display one density chart for that feature, with two red lines indicating the feature values from the trajectory pair.

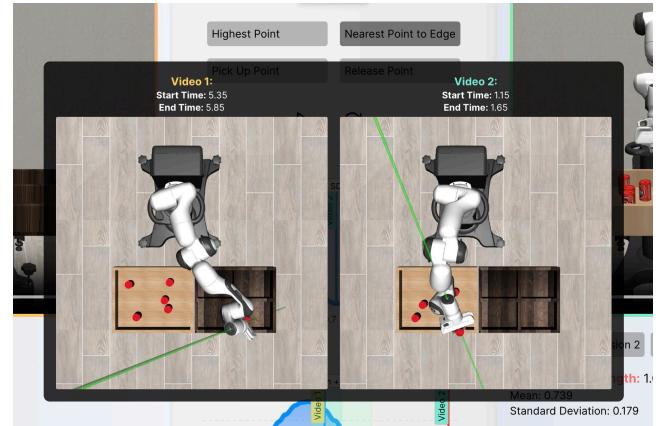
**4.3.3 Progress stepper and feedback (*DF4*).** Finally, we add a progress stepper (Figure 4) at the bottom of the interface, as well as prompting windows (Figures 7 and 8) to provide users with feedback regarding progress. The progress bar does not explicitly state the number of completed trajectory pairs to prevent overwhelming the labelers with a large number of pairs. Instead, it updates to the next step before each attention-checking pair and provides feedback messages from the server to positively reinforce the user's progress. These components fulfill *DR3* of real-time attention monitoring and feedback provision from the user interface.



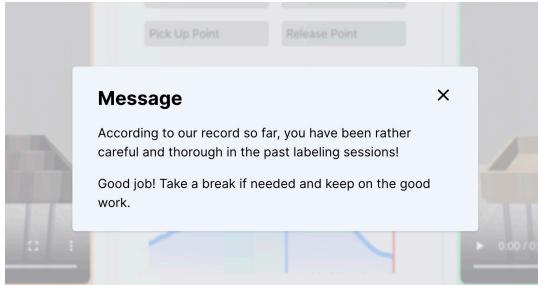
**Figure 4:** The primary component of the user interface is two juxtaposed trajectory videos. Labelers can play the videos simultaneously using the play button in the middle or individually using separate controls. The labeler can click one button on the top to indicate a preference. Auxiliary features include looping videos with *keyframe buttons* (Section 4.3.1, Figures 5 and 6), outlying *feature value distribution* (Section 4.3.2), and a *progress stepper* with messages prompting at each step (Section 4.3.3, Figures 7 and 8).



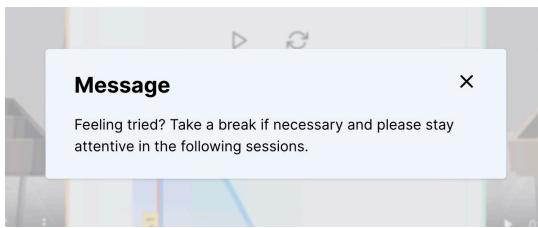
**Figure 5: Collision keyframe preview when hovering on the buttons.**



**Figure 6: Common keyframe preview when hovering on the buttons.**



**Figure 7: Encouraging messages.**



**Figure 8: Rest reminder messages.**

## 5 EVALUATION

We conducted a between-subjects study to evaluate FARPLS's effectiveness in improving the quality of preference labels with the Institutional Review Board (IRB) approval from the University Research Ethics Committee.

### 5.1 Participants

Forty-two participants (21 identified themselves as female, 18 identified themselves as male, and 2 identified themselves as non-binary, in the age range of 19 to 40 with a mean of 23.4 and a standard deviation of 3.2), who did not participate in the formative study, were recruited in the user study. Their familiarity with the data labeling system (1 for Not familiar at all – 5 for Extremely familiar) ranges from 1 to 4 with a mean of 2.2 and a standard deviation of 0.9. Additionally, their familiarity with robotics (1 for Not familiar at all – 5 for Extremely familiar) ranges from 1 to 4 with a mean of 2.0 and a standard deviation of 0.8.

### 5.2 Task and Condition

We deployed two online trajectory preference labeling systems to a web server for the user study: a) the baseline system with a conventional interface with two videos side-by-side and the stepper only (Figure 9a), and b) our proposed system FARPLS with features and keyframes shown as auxiliary information additional to the two videos and the stepper (Figure 9b). The method of between-subjects design was used, and we randomly split the participants into two groups, with baseline group ( $N = 21$ ) and FARPLS group ( $N = 21$ ).

To prepare data with similar feature distributions with *PickPlaceCans*, we calculated sample weights based on the feature values extracted from each trajectory and stratified sampled 30 representative samples (i.e.,  $30 \times 29/2 = 435$  unique trajectory pairs) from

*PickPlaceCans* with the sample weights for the user study. Our comparative study utilizes these 30 sampled trajectories to compare the performance of the two systems. We invited each participant to label 115 pairs (with 105 unique pairs from the sampled trajectories and 10 pairs for consistency checking) of robot pick-and-place trajectories through the system. In the baseline condition, the system randomly assigns 105 unique pairs to each participant, at the same time, making sure that at least 5 participants label each pair. In the FARPLS condition, FARPLS first clusters the 30 trajectories using the criterion vector series and dynamically prompts the trajectory pairs according to the strategy mentioned in Section 4.2.1, which balances different ranking metrics when guaranteeing label skewness.

Labeling 105 unique pairs consists of 10 steps, each step containing 10 (the number is 15 for the first step) unique pairs and 1 consistency checking pair. At the end of each step, both systems randomly select a consistency-checking pair from the labeled pairs and prompt it to the user. We did not tell the participants the numbers of unique pairs and consistency-checking pairs. Still, both systems can show the status (incomplete/active/completed) of 10 steps in the stepper to the participants and prompt a stop message when all pairs are labeled.

### 5.3 Procedure

First, the host of the study session provided instructions on the robot task, the labeling system for the corresponding group, and the labeling task and ensured that participants were clear about everything. Then, the participants were asked to label 115 pairs of robot pick-and-place trajectories through the system. Finally, we conducted a post-study survey with the 7-point Likert scale and open-ended interview questions. The FARPLS group was administered additional 7-point Likert scale questions regarding the auxiliary features in FARPLS, while distinct open-ended interview questions were posed to the two groups. The whole study session took about 90 to 120 minutes, varying depending on the learning, loading, and labeling speed.

### 5.4 Evaluation Measurement

We collected quantitative measures by the collected preference data, logging participant interactions, and Likert scale questions. We also conducted a semi-structured interview to collect participants' other feedback. Table 4 lists the evaluation metrics used in our study. The objective metrics include consistency and labeling time. The subjective metrics include cognitive load, confidence and challenges in Section 3.3. For the FARPLS group, we further explored their perspectives on the auxiliary information provided by the system and their perceptions of how well the design goals were fulfilled (Table 5).

Additionally, we conducted a semi-structured interview to collect participants' other feedback. For both groups, we asked about participants' criteria with features and priorities and their comments and suggestions on the system. For the baseline group, we also discussed the features and keyframes in Table 1 at the end. For the FARPLS group, we asked more about their opinions on each design requirement (Section 3.4).

**Table 4: Evaluation Metrics for the User Study, including Objective and Subjective Metrics.**

Objective Metrics	Definitions
Consistency	The percentage of consistent labels in the 10 consistency checking pairs, taking interval values in $\{0, 0.1, \dots, 0.9, 1\}$
Labeling Time	The total labeling time (loading time eliminated), which takes continuous values in $[0, \infty)$
Subjective Metrics	7-Likert Scale Questions
	General Questions
Cognitive Load	How mentally challenging was it for you to compare and specify your preference over the video pairs in general? (1 - “Not mentally challenging at all” to 7 - “Extremely mental challenging”)
Confidence	How confident were you about your labels in general? (1 - “Not confident at all” to 7 - “Extremely confident”)
	Questions About Three Challenges
Please rate how much you agree with the following statements on a scale of 1 to 7, where 1 is “strongly disagree” and 7 is “strongly agree”:	
C1 Comparison Criteria	<b>C1 [criteria establishment]</b> I can easily establish comparison criteria in general.
	<b>C1-1 [criteria coverage]</b> My comparison criteria can cover all the new situations in the later videos.
	<b>C1-2 [feature coverage]</b> I am clear about the set of features I rely on to decide the priority of each criterion.
	<b>C1-3 [feature distribution]</b> I am clear about the scope of each feature required to determine the priority of each criterion.
C2 Trajectory Details	<b>C2 [detail overlooking]</b> I feel that I may overlook some important details when viewing and comparing two trajectories.
	<b>C2-1 [robotic knowledge]</b> My knowledge of this robot arm task affects the kind of details I pay attention to.
	<b>C2-2 [feature support]</b> The system provides enough support for me to identify features that may be important to this robot arm task in practice.
	<b>C2-3 [comparison support]</b> The system provides enough support for me to compare the differences between the two videos.
C3 User Experience	<b>C3-1</b> The preference labeling process is <b>[easy]</b> .
	<b>C3-2</b> The preference labeling process is <b>[boring]</b> .
	<b>C3-3</b> I receive <b>[encouragement]</b> in the preference labeling process.
	<b>C3-4</b> I receive <b>[feedback]</b> on my performance in the preference labeling process.
	<b>C3-5</b> I find the preference labeling process <b>[rewarding]</b> .

## 6 RESULTS

To answer RQ3, we comprehensively analyze all the metrics for each group. We first conduct significant tests for each metric in two-sided, and then for significant metrics, we conduct additional one-sided tests. We also answer how well FARPLS solves the challenges by each design requirement and the participants’ feedback on the design requirements. We further highlight the participants’ most insightful comments and suggestions.

### 6.1 Consistency

The FARPLS group has a significantly higher consistency score than the baseline group according to the Mann-Whitney U test [53],

$U = 321.5, p = 0.0046^{**} < 0.01$ . The participants in the baseline group have an average consistency score of 0.79 ( $SD = 0.15$ ), and the participants in the FARPLS group have an average consistency score of 0.89 ( $SD = 0.16$ ). Figure 10a shows the distribution of consistency scores in boxplots for both groups. The result indicates that participants are more consistent with the help of FARPLS.

### 6.2 Labeling Time

**6.2.1 Total Labeling Time.** Participants spend significantly more time labeling with FARPLS than with the baseline system according to Welch’s t-test,  $t(36.18) = 3.127, p = 0.0017^{**} < 0.01$ , 95% CI  $[242.36, \infty]$ . On average, the participant in the baseline group spent 1350s ( $SD = 448$ ) labeling all the trajectories, while

**Table 5: Subjective Questions for FARPLS Group Only**

Metrics	7-Likert Scale Questions
Please rate how much you agree with the following statements about the auxiliary information provided by the system on a scale of 1 to 7, where 1 is “strongly disagree” and 7 is “strongly agree”:	
Auxiliary Information	<p><b>AX1</b> The auxiliary information provided by the system is [<b>informative</b>].</p> <p><b>AX2</b> The auxiliary information provided by the system is [<b>relevant</b>].</p> <p><b>AX3</b> The auxiliary information provided by the system is helpful for establishing preference [<b>criteria</b>].</p> <p><b>AX4</b> The auxiliary information provided by the system prevents me from observing more [<b>details</b>] in the videos.</p> <p><b>AX5</b> The auxiliary information provided by the system is helpful for [<b>comparing</b>] the differences between the two videos.</p> <p><b>AX6</b> The auxiliary information provided by the system is [<b>overwhelming</b>].</p> <p><b>AX7</b> The auxiliary information provided by the system is [<b>distracting</b>].</p>
We will present you with three designs of this system. Please rate how helpful each design is in a particular aspect on a scale of 1 to 7, where 1 is “not at all helpful” and 7 is “very helpful”:	
1. We group the trajectories according to a set of features illustrated in Table 1. Subsequently, we present the trajectories with greater variations in these features to you initially.	
DR1 Prompting Strategy	<p><b>DR1C1-1 [initial familiarity]</b> How helpful is this design in facilitating familiarity with diverse situations and establishing criteria for comparison?</p> <p><b>DR1C2-2 [detail aware]</b> How helpful is this design in improving your perception of the details of the trajectory, regardless of your level of robotics expertise?</p>
2. This system presents the distributions of representative trajectory features and keyframes from a particular view.	
DR2 Feature and Keyframe	<p><b>DR2C1-2 [criteria priority]</b> How helpful is this design in assisting your comprehension of each feature’s range to determine the priority of your criteria?</p> <p><b>DR2C2 [sense making]</b> How helpful is this design in enhancing your understanding of this robot task and gaining access to more information?</p>
3. This system offers a real-time attention-monitoring feature. If you label inconsistently, the system will prompt you to take a break. On the other hand, if you label consistently, it will motivate you to continue.	
DR3 Attention Monitoring and Feedback	<p><b>DR3C3-2</b> How helpful is this design in [<b>decreasing the boredom</b>] of labeling preferences?</p> <p><b>DR3C3-5</b> How helpful is this design in [<b>increasing the reward</b>] of the preference labeling process?</p>

the participant in the FARPLS group spent 1877s ( $SD = 628$ ). Figure 10b shows the distribution of total labeling time spent by two groups of participants in boxplots. The result is within expectations, since participants in the FARPLS group need to spend more time observing the auxiliary information in FARPLS.

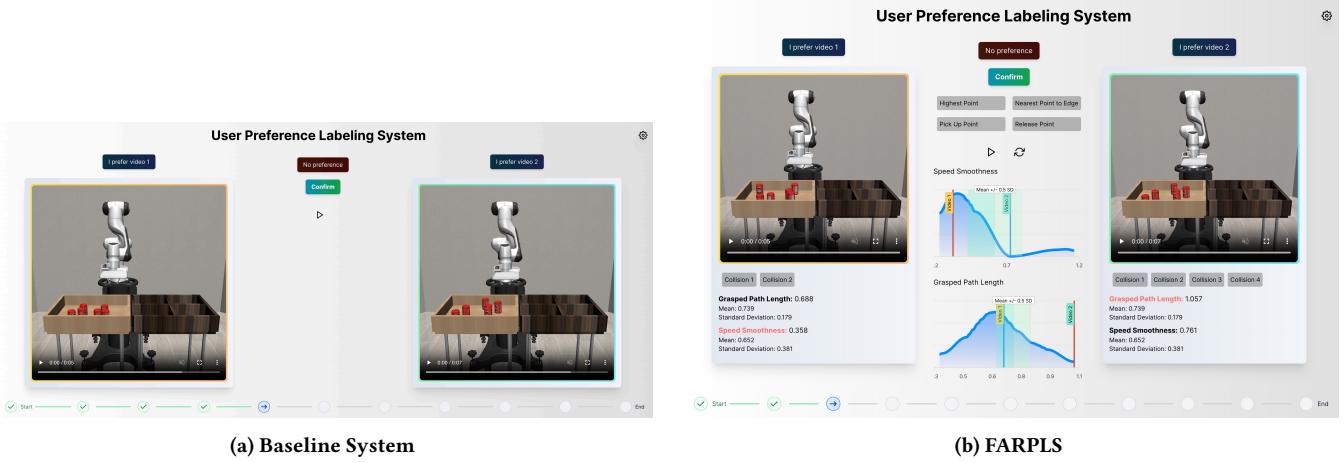
**6.2.2 Learning Curve.** We plot a learning curve to analyze how the participants’ labeling time for each trajectory pair changes during the experiment in Figure 11. The fitted smoothing lines (via Locally Weighted Scatterplot Smoothing (LOESS) method [19]) show the trend of the average labeling time per pair of participants. The average labeling time per pair of participants decreases as the experiment progresses, which indicates that the participants are getting more familiar with the labeling task and their comparison criteria. A sharper decreasing trend in the FARPLS group than in the baseline group also indicates that the participants in the FARPLS group are learning faster than the participants in the baseline group. There

is also a slight increase at the end of the FARPLS group’s learning curve, possibly due to the prompted pairs’ increased difficulty.

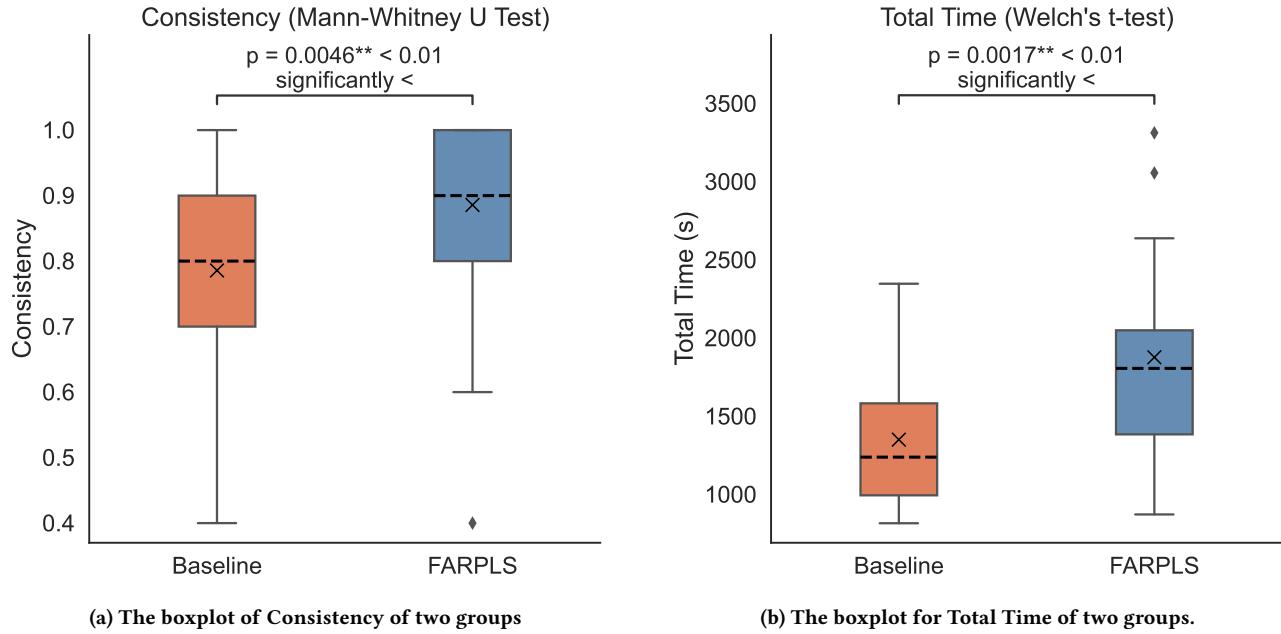
### 6.3 Cognitive Load and Confidence

Figure 12 shows the distributions of the ratings of two subjective metrics, Cognitive Load and Confidence.

**6.3.1 Cognitive Load.** There is no significant difference in the cognitive load scores between the two groups according to the Mann-Whitney U test,  $U = 190.0$ ,  $p = 0.4375 > 0.05$ . The participants in the baseline group have an average cognitive load score of 3.28 ( $SD = 1.42$ ), and the participants in the FARPLS group have an average cognitive load score of 3.00 ( $SD = 1.64$ ). This result indicates that the additional design features in FARPLS do not affect the participants’ cognitive load.



**Figure 9: System interfaces of two evaluation conditions: Baseline System and FARPLS. The Baseline System features two side-by-side videos and a stepped progress bar, while FARPLS incorporates additional information, including keyframes, feature distributions, and feedback messages.**



**Figure 10: The distributions of two subjective metrics Consistency and Total Time in boxplots comparing two system conditions.**

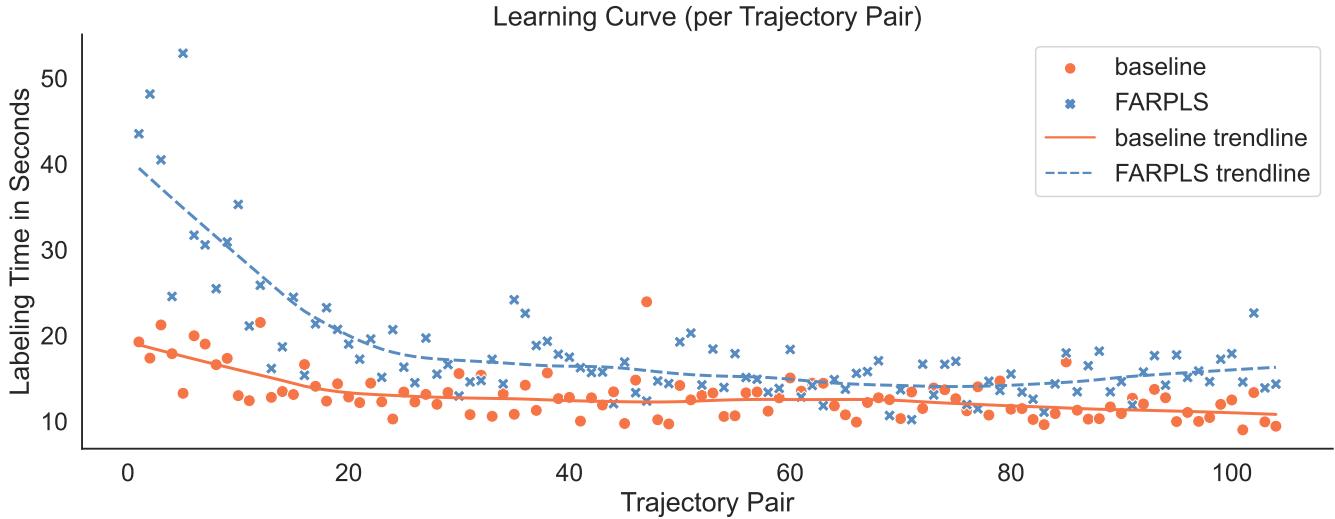
**6.3.2 Confidence.** There is no significant difference in the confidence scores between the two groups according to the Mann-Whitney U test,  $U = 166.0$ ,  $p = 0.2027 > 0.05$ . The participants in the baseline group have an average confidence score of 4.81 ( $SD = 1.40$ ), and the participants in the FARPLS group have an average confidence score of 5.29 ( $SD = 1.16$ ). The participants are generally confident in their labeling according to the means ( $> 4$ ) and the plots in Figure 12. However, the baseline group's Consistency metric is significantly lower than the FARPLS group. This

finding suggests that participants may have overconfidence in their self-assessments of the Confidence metric.

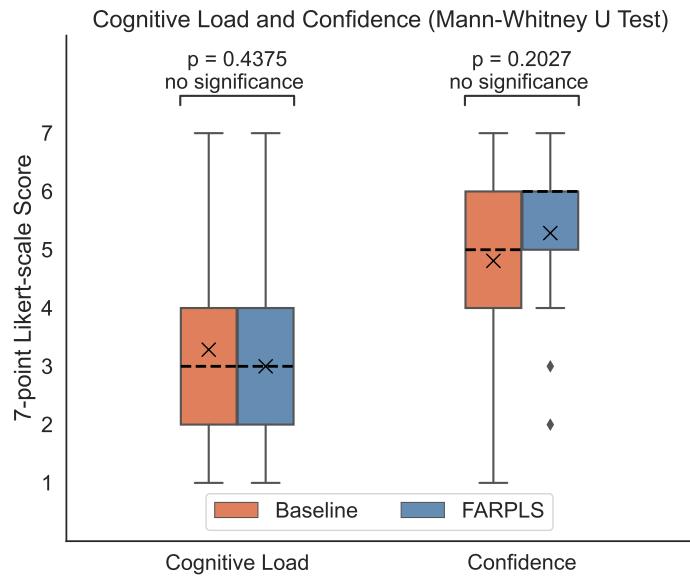
#### 6.4 Challenges and Design Requirements

Table 6 shows the statistics for metrics and Figures 13 to 15 show the distributions of the metrics.

**6.4.1 C1 - Difficulty in forming criteria.** Figure 13 shows the distributions of metrics about *C1 - Difficulty in forming criteria* in boxplots comparing two system conditions. According to the statistics in Table 6, we have the following results:



**Figure 11:** Learning curve showing the participants' average labeling time per pair.



**Figure 12:** The distributions of Cognitive Load and Confidence in boxplots comparing two system conditions.

*C1 [criteria establishment].* This metric in FARPLS group is significantly higher than that in the baseline group according to the Mann-Whitney U test. Participants in the FARPLS group find it significantly easier to establish criteria in general compared to the baseline group.

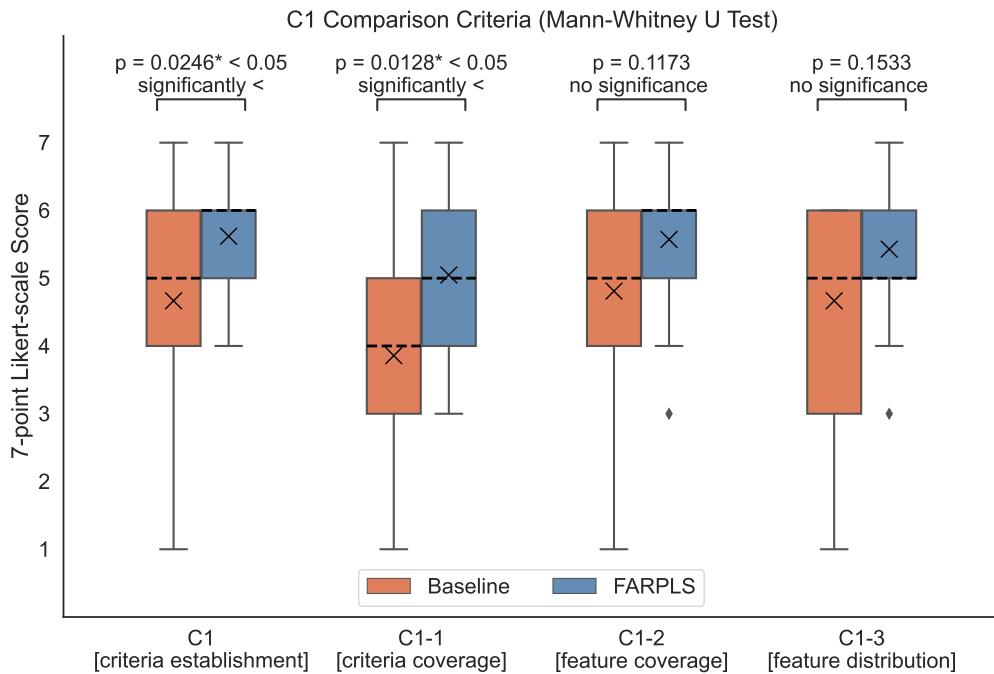
*C1-1 [criteria coverage].* This metric in FARPLS group is significantly higher than that in the baseline group according to the Mann-Whitney U test. Participants in the FARPLS group reckon their criteria can cover more new situations in the later videos than those in the baseline group.

*C1-2 [feature coverage].* This metric has no significant difference between the two groups according to the Mann-Whitney U test. The results show that participants in both groups reckon they are clear about the features to decide the priority of each criterion.

*C1-3 [feature distribution].* This metric has no significant difference between the two groups according to the Mann-Whitney U test. The results show that participants in both groups reckon they are clear about the scope of features to determine the priority of each criterion.

**Table 6: Challenge Metrics Statistics and Mann-Whitney U Test Results**

Metrics		Baseline		FARPLS		Mann-Whitney U Test (FARPLSv.s. Baseline)		
		mean	std	mean	std	U-value	alt. hypo.	p-value
C1	[criteria establishment]	4.67	1.59	5.62	0.86	295.5	greater	0.0246*
C1-1	[criteria coverage]	3.86	1.77	5.05	1.20	307.5	greater	0.0128*
C1-2	[feature coverage]	4.81	1.54	5.57	1.03	281.0	two-sided	0.1173
C1-3	[feature distribution]	4.67	1.46	5.43	0.98	275.0	two-sided	0.1533
C2	[detail overlooking]	4.00	1.70	3.90	1.41	214.5	two-sided	0.8870
C2-1	[robotic knowledge]	3.81	1.91	3.43	1.80	193.0	two-sided	0.4901
C2-2	[feature support]	4.10	1.73	5.76	0.94	349.0	greater	0.0005***
C2-3	[comparison support]	4.76	1.76	5.86	1.20	306.5	greater	0.0125*
C3-1	[easy]	4.71	1.95	5.33	1.43	256.0	two-sided	0.3643
C3-2	[boring]	4.95	1.63	4.19	1.36	152.5	less	0.0419*
C3-3	[encouragement]	3.62	1.60	4.57	1.54	288.0	greater	0.0424*
C3-4	[feedback]	2.95	1.56	4.81	2.02	334.0	greater	0.0019**
C3-5	[rewarding]	3.48	1.83	4.71	1.23	312.0	greater	0.0101*

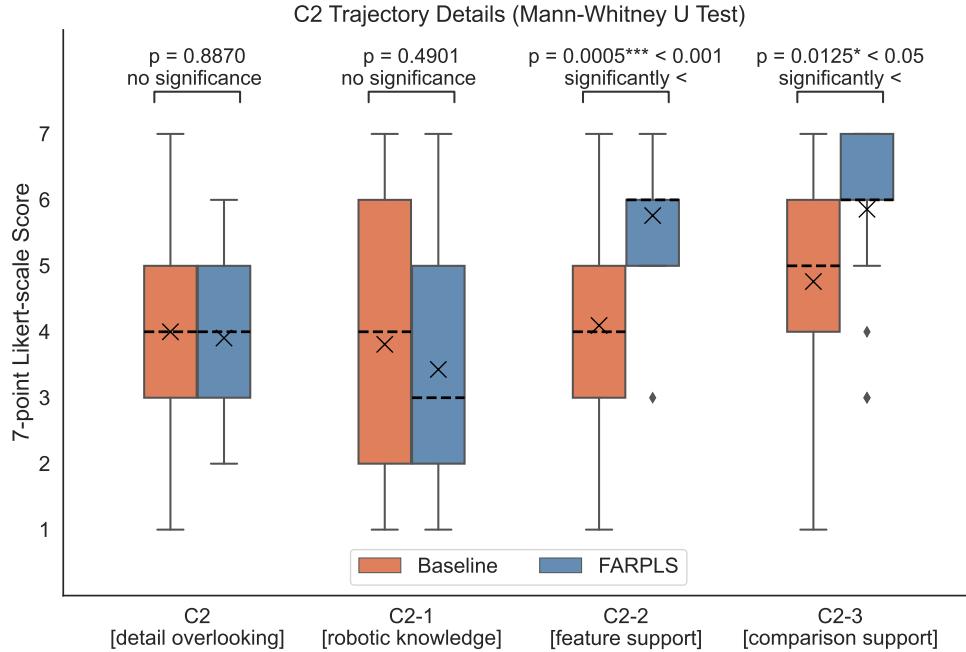
**Figure 13: The distributions of metrics about C1 (Comparison Criteria) in boxplots comparing two system conditions.**

*Summary.* The significant results in C1 and C1-1 (Table 6) show that FARPLS can successfully mitigate the challenges in forming criteria.

**6.4.2 C2 - Overlooking trajectory details.** Figure 14 shows the distributions of metrics about C2 - Overlooking trajectory details in boxplots comparing two system conditions. According to the statistics in Table 6, we have the following results:

**C2 [detail overlooking].** This metric has no significant difference between the two groups according to the Mann-Whitney U test. The neutral results of the metric C2 in both groups (both means  $\sim 4.00$  and both medians = 4) indicate that people do not know whether they overlook trajectory details.

**C2-1 [robotic knowledge].** This metric has no significant difference between the two groups according to the Mann-Whitney U test. Considering that the two groups' mean metric scores for C2-1



**Figure 14: The distributions of metrics about C2 (Trajectory Details) in boxplots comparing two system conditions.**

(3.81 and 3.43) are  $< 4.00$ , we can infer that although with large disagreements ( $SD = 1.91$  and  $= 1.80$ ), many participants in both groups think their knowledge of robotics does not affect their labeling process.

**C2-2 [feature support].** This metric in FARPLS group is significantly higher than that in the baseline group according to the Mann-Whitney U test,  $U = 349.0$ ,  $p = 0.0005^{***} < 0.001$ . Participants in the FARPLS group feel significantly more support than those in the baseline group to identify features that may be important to the robot arm task in practice.

**C2-3 [comparison support].** This metric in FARPLS group is significantly higher than that in the baseline group according to the Mann-Whitney U test. Participants in the FARPLS group feel significantly more support than those in the baseline group to compare the differences between the two videos.

**Summary.** The significant improvements in C2-2 and C2-3 (Table 6) show that FARPLS can successfully assist labelers in identifying more features and easily eliciting their preferences.

**6.4.3 C3 - Difficulty in maintaining focus.** Figure 15 shows the distributions of metrics about C3 - Difficulty in maintaining focus in boxplots comparing two system conditions. According to the statistics in Table 6, we have the following results:

**C3-1 [easy].** This metric has no significant difference between the two groups according to the Mann-Whitney U test. The scores of C3-1 are high in means ( $4.71$  and  $5.33 > 4.00$ ) and medians ( $5$  and  $6 > 4$ ) for both groups. The results show that participants in both groups reckon the labeling task easy.

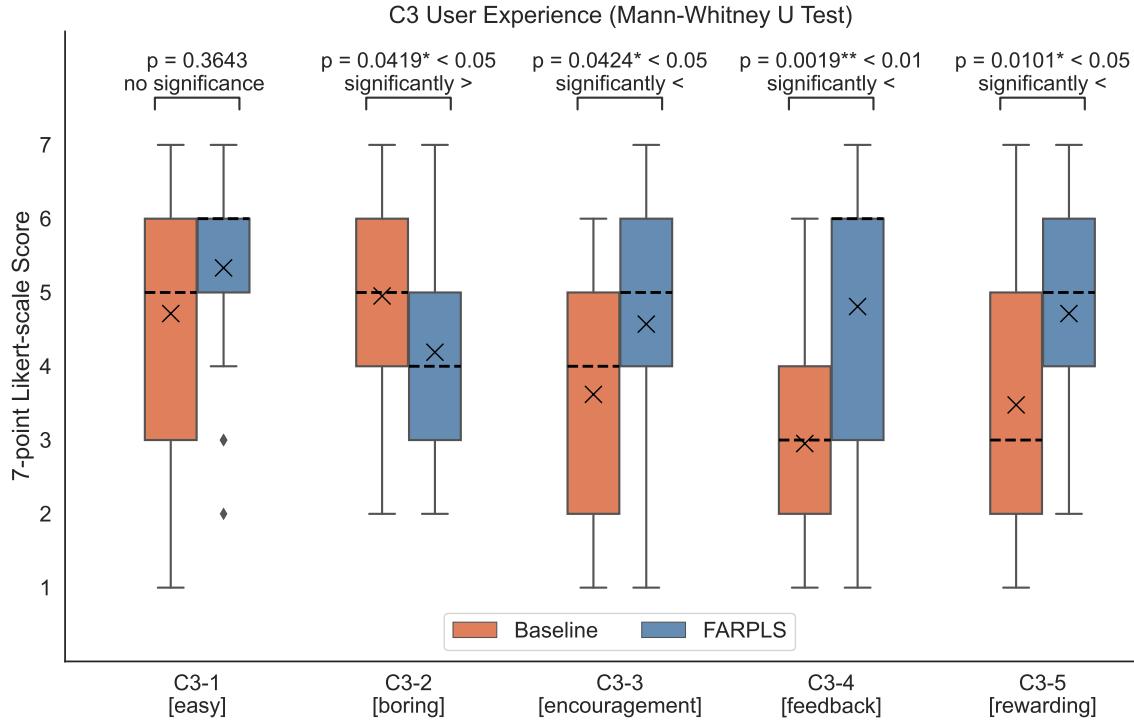
**C3-2 [boring].** This metric in FARPLS group is significantly lower than that in the baseline group according to the Mann-Whitney U test. Participants in the baseline group find the labeling process more boring than those in the FARPLS group.

**C3-3 [encouragement].** This metric in FARPLS group is significantly higher than that in the baseline group according to the Mann-Whitney U test. Participants in the FARPLS group reckon they receive more encouragement in the labeling task than those in the baseline group.

**C3-4 [feedback].** This metric in FARPLS group is significantly higher than that in the baseline group according to the Mann-Whitney U test. Participants in the FARPLS group think they receive more feedback on their labeling performance than those in the baseline group.

**C3-5 [rewarding].** This metric in FARPLS group is significantly higher than that in the baseline group according to the Mann-Whitney U test. Participants in the FARPLS group feel the labeling process more rewarding than those in the baseline group.

**Summary.** The significant improvements (Table 6) in C3-2, C3-3, C3-4 and C3-5 show that the participants in the FARPLS group find the labeling process less tedious and more engaging than those in the baseline group. The insignificant result of C3-1 [easy] indicates that the participants think the labeling task is easy, resulting in their confidence in their criteria (aligned with the Confidence metric) and their belief in the comprehensiveness of their feature criteria and their familiarity with the distribution of features (aligned with the metrics C1-2 [feature coverage] and C1-3 [feature distribution]). As a result, participants do not think their robotic knowledge affects



**Figure 15: The distributions of metrics about C3 (User Experience) in boxplots comparing two system conditions.**

the details they pay attention to (aligned with the metric C2-1 [robotic knowledge]). Therefore, the metrics C3-1, C1-2, C1-3, and C2-1 are logically aligned with Confidence. We will discuss the implications of this finding in the discussion section.

**6.4.4 Design requirements.** The additional metrics for the FARPLS group are for participants to rate the auxiliary information and the helpfulness of design requirements of FARPLS, reported in Table 7 and Figure 16. As we can see from the results (relevant means and medians comparing with 4), the participants find the auxiliary information informative, relevant, and helpful for establishing comparison criteria. Moreover, the auxiliary information does not prevent the participants from observing more details in the videos and is helpful for them to compare the trajectories, neither overwhelming nor distracting.

The rest questions are designed to check whether the design requirements of FARPLS solve the challenges mentioned in Section 3. We dynamically arranged the order to balance labeling difficulty in the case of DR1. Our prompting strategy during the initial stage significantly facilitated participants in quickly gaining familiarity with the labeling task and establishing their comparison criteria (DR1C1-1,  $M = 6.24$ ). Furthermore, it enabled them to comprehend the intricate details of the trajectories (DR1C2-2,  $M = 6.24$ ). For DR2, the participants find the design helpful for them to understand the feature value distributions to determine the priorities (DR2C1-2,  $M = 5.62$ ) and access more information to make sense of the robot task (DR2C2,  $M = 5.90$ ). As shown in the metric results of DR3, the participants find the design moderately helpful in decreasing the boredom (DR3C3-2,  $M = 5.05$ ) and increasing the reward (DR3C3-5,

$M = 5.05$ ). Therefore, participants find all the design requirements of FARPLS helpful for the corresponding challenges connected to the DRs in Figure 2.

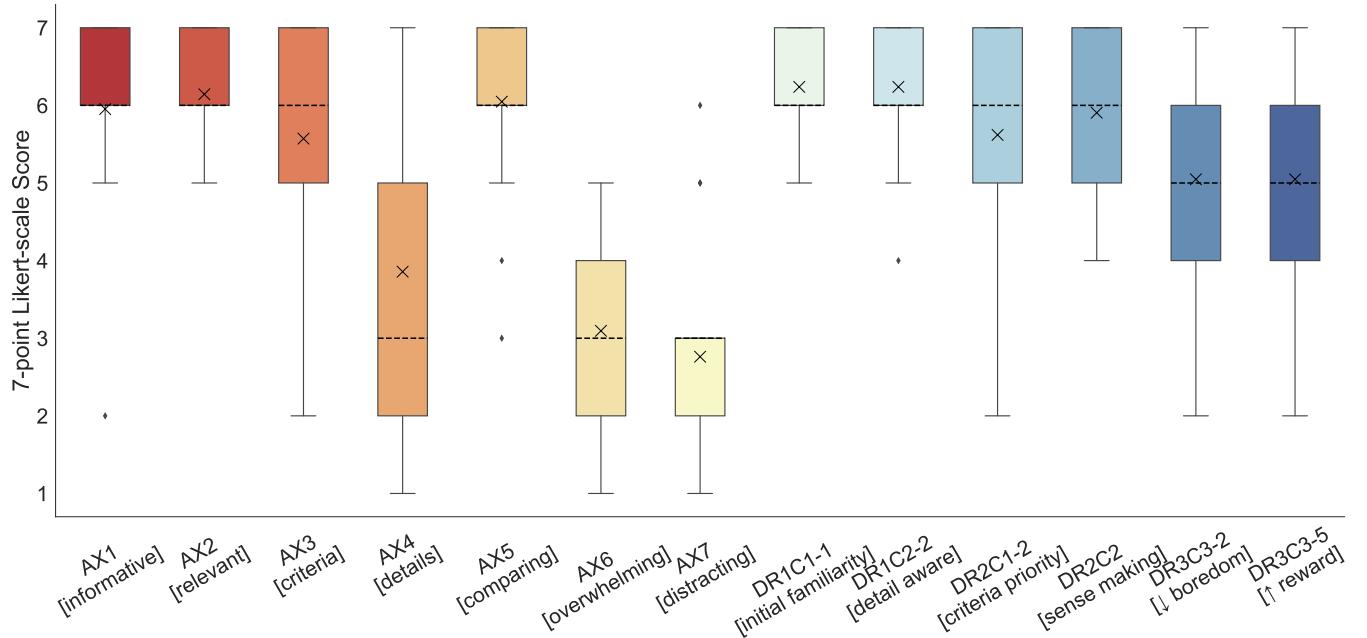
## 6.5 Semi-structured Interview

**6.5.1 Features in participants' comparison criteria.** We compared the number of features noticed by participants from two groups. We clarified subtle distinctions in our feature categorization to the participants and tallied their explicit indications of whether they noticed each of them. “Power Usage”, “Contact Force”, and “Orientation” are the three least noticed features with 18, 14, and 8 participants out of 21 neglecting them, respectively, as is shown in Table 8. In the FARPLS group, we integrated all 11 features into the system. Participants reported that they noticed these features and were free to assign different priorities to each. Thus, they take more time observing auxiliary information than the baseline group, which can be a reason for the significantly increased total labeling time.

According to the participants, the improved system provided features in advance and thus broadened their perspectives and encouraged further elicitation. It also revealed subtle differences challenging or impossible to discern with the naked eye, aligned with *C2 - Overlooking trajectory details*. Though participants tended to set lower priority to some of the features, these features were crucial, especially in cases where their primary criteria exhibited limited variation and participants needed to turn to more features for guidance. For example, one user stated that he would rely on

**Table 7: Statistics of additional metrics about auxiliary information and design requirements of the FARPLS group.**

	AX1 [informative]	AX2 [relevant]	AX3 [criteria]	AX4 [details]	AX5 [comparing]	AX6 [overwhelming]	AX7 [distracting]
mean	5.95	6.14	5.57	3.86	6.05	3.10	2.76
std	1.12	0.73	1.40	2.13	1.12	1.41	1.30
	DR1 Prompting Strategy		DR2 Feature and Keyframe		DR3 Attention Monitoring and Feedback		
	DR1C1-1	DR1C2-2	DR2C1-2	DR2C2	DR3C3-2	DR3C3-5	
mean	6.24	6.24	5.62	5.90	5.05	5.05	
std	0.70	0.77	1.50	0.94	1.47	1.53	

**Figure 16: Distributions of additional metrics about auxiliary information and design requirements of the FARPLS group in boxplots.**

lower-priority features like “Power Usage” when the primary criteria (e.g., “Collision”, “Trajectory Smoothness”) cannot distinguish the trajectories.

**6.5.2 Other comments and suggestions.** Participants provided various comments and suggestions regarding the system and the labeling process.

In terms of the labeling order and process, participants encountered certain pairs that were challenging to compare, possibly those where *User Familiarity*, *Pair disagreement* or *Cluster Disagreement* dominated other metrics in Table 3, and some easier pairs, possibly those where *Cluster Coverage* or *Pair Similarity* dominated other metrics in Table 3. Trajectory pairs with different difficulty levels are interspersed in the labeling process. Therefore, the participants found the labeling process’s workload more manageable and experienced increased engagement in the task. Several participants mentioned that they would feel more rewarded if they could see

the improvements in the robot’s performance during the labeling process.

From the features and keyframes perspective, some participants recommended that the system should always highlight unobservable features and provide more detailed explanations or definitions of the trajectory features, particularly for those unfamiliar with the robotic task. Moreover, participants wished to view the complete list of feature distributions and keyframes and select the ones to display and compare. Some participants suggested that the system could provide more detailed written descriptions instead of letting the host explain the trajectory features orally. One participant suggested that the system should provide a way to quantify the collision’s severity since some collisions do not affect the safety of the object and the robot; in contrast, some collisions may cause severe damage. A possible way is to use the collision force to quantify the severity of each collision.

**Table 8: Number of baseline group participants overlooking each feature.**

Feature	Number of Participants
Collision	0
Distance	2
Contact Force	14
Speed	7
Path Length	6
Time	6
Power Usage	18
Speed Smoothness	7
Trajectory Smoothness	2
Orientation	8
Grasp Position	7

From the workload and engagement perspective, although more engaged using FARPLS, many participants still feel overwhelmed by the number of trajectory pairs for the labeling workload and hope to have fewer pairs to label. One insightful suggestion is leveraging gamification to make the labeling process more engaging by providing more variety in the task scenario presented.

## 7 DISCUSSION

Our user study shows that the participants in the FARPLS condition have significantly higher labeling consistency than those in the baseline condition. Although the total labeling time of the FARPLS group is significantly longer than that of the baseline group, the participants in the FARPLS group find it significantly easier to form comparison criteria and maintain engagement than those in the baseline group. They also consider FARPLS helpful and easy to use. According to participant ratings, our proposed design requirements of FARPLS help tackle the corresponding challenges identified in the formative study. In the following subsections, we discuss the potential research implications derived from the participants' feedback, the room for improvement of each component, the generalizability of the proposed tool FARPLS, and the possible limitations of this work.

### 7.1 Implications for Future Research

**7.1.1 Support for sense-making and preference elicitation.** Our dataset *PickPlaceCans* augments conventional robot task data with descriptive statistics of a wide variety of trajectory features and keyframes of critical moments in task videos. We leverage such information to cluster robot trajectories, prompt labeling arrangement, and facilitate comparison on the FARPLS interface. The results (Sections 6.4.1, 6.4.2 and 6.4.4) show that these features and keyframes can help users better understand complex trajectories, establish preference criteria, and make decisions. Through the keyframes, users can visually identify the differences in a specific feature between the trajectories through a quick glance at the related video segment. For example, users can compare the height at which the robot drop the object by the corresponding

juxtaposed frames from the two videos. With the plots of feature value distributions, users can determine the priority of the features.

Due to the space limitation, in our current design, FARPLS chooses to present the distribution of a feature on the labeling interface if its value in any of the given two trajectories is considered an outlier. The assumption is that such features are distinctive and worthy of attention. Another possible way to select features for display is to let users specify features of interest, as some participants mentioned in the interview. However, users may not know what features matter to them, especially at the beginning, or their judgment is biased. In other words, both the adaptive and the adaptable approaches have their own advantages and disadvantages. Future work may further explore the usability and user-friendliness of a mixed-initiative approach [12], ensuring user agency while trying to mitigate individual biases. Overall, these features serve as an abstraction of the trajectories to boost comprehension. We only employ familiar basic charts and keyframes to illustrate the features in this work. Future research can explore alternative means to present such auxiliary information on the side or directly overlay on top of a task trajectory. For example, a potential direction can be utilizing situated visualization in augmented reality (AR) [15] for preference elicitation of trajectories.

**7.1.2 Overconfidence in self-assessments.** Our study finds that the metrics related to participants' self-assessments (C3-1 [easy], Confidence, C1-2 [feature coverage], C1-3 [feature distribution], and C2-1 [robotic knowledge]) did not show significant differences between the FARPLS and the baseline groups. However, we observe a significant improvement in the Consistency metric of the FARPLS group. Moreover, during the interviews, many baseline group participants admitted that they failed to notice some important features (Section 6.5.1). These results suggest that the participants may be overconfident in their preference labels, known as *Dunning-Kruger effect* [44], even though most claimed to be unfamiliar with robotics. This finding aligns with previous studies [27, 31] that have identified the overconfidence bias in crowdsourcing tasks. In such tasks, labelers with low ability tend to overestimate their performance, leading to insufficient observations and inconsistent labels. This insight highlights the importance of assisting labelers in calibrating their confidence. Possible methods include but are not limited to training the non-expert labelers [75] to improve their expertise, increase labelers' awareness of cognitive biases by providing trial-by-trial feedback [36], and introducing labeler collaboration in the preference elicitation process [78]. Future research can investigate the efficacy of various confidence calibration mechanisms and the consequent effect on label quality.

**7.1.3 Design of prompting mechanisms.** FARPLS tries to balance the model's information collection and the user's experience by considering the similarities between trajectories, disagreements among labelers, and individual familiarity with the trajectory pair when dynamically adjusting the comparison order. The results from the FARPLS group suggest that presenting trajectory pairs that cover more distinct cases at the initial stage can improve their ability to observe a wide variety of details and establish relatively reliable criteria (Section 6.4.4). Additionally, according to the interview, the FARPLS's adaptive prompting order helps users stay engaged in the task and promotes a positive user experience. Nevertheless,

how each component (e.g., pair similarity, user familiarity, and disagreements) impacts the performance and experience is yet to be assessed. Future work can conduct ablation studies to determine the effect of individual components, comparing the effectiveness of the robot reward models trained on the labeling results. Furthermore, future studies can experiment with other factors and mechanisms to achieve an optimal balance between the model's uncertainty, users' experience, and labeling quality. For example, inspired by active learning [3, 38, 71], researchers could add a model informative dimension to determine the priority of comparative pairs. They may also incorporate annotation curricula [48] or gamification [70] in the labeling process to reduce labelers' learning curve and improve their engagement.

## 7.2 Generalizability of FARPLS

Our study demonstrates the effectiveness of our pairwise preference labeling system, FARPLS, in improving the consistency of human feedback in the pick-and-place task. This subsection discusses the generalizability of FARPLS.

**7.2.1 Generalizing to other tasks.** FARPLS can also be generalized to other robot arm manipulation tasks besides the pick-and-place task. In the formative study, we interviewed the participants about the issues they may encounter in labeling other robot tasks. They reflected that the three challenges identified in Section 3.3 exist regardless of the tasks. They further commented that the criteria and trajectory features of interests are likely to be similar across different robot arm manipulation tasks, but the priority of individual criteria may vary in different scenarios. That is to say, the criteria, features, challenges, and design requirements we summarized from the formative study are not limited to the pick-and-place task. To tailor our design to other tasks, it is only necessary to redefine the list and the formulas of trajectory features and extract the corresponding keyframes according to the specific requirements of the new task. For example, the *Distance* feature can be defined as the distance of the pouring point to the cup in the pouring task. Future work can further investigate the effectiveness of FARPLS adapting to other robot arm manipulation tasks. For example, how well FARPLS could be adapted to more complex tasks involving multiple objects, multiple parties, or multiple steps, such as the kitchen tasks in [34].

**7.2.2 Generalizing to other feedback types.** Since the highlights of our design for FARPLS are the augmentation of the features and prompting strategies, it is flexible enough to accommodate various types of human feedback, including but not limited to pairwise comparison [33, 46, 63], multiple ranking [9, 10, 60, 89], rating [13], etc., as long as the input data are consistent. The dataset *PickPlace-Cans* and the ranking metrics in Table 3 are prepared for individual trajectories on the server side, regardless of how the trajectories are presented and annotated on the client side. To generalize FARPLS to other feedback types, we only need to adjust the user interface and the prompting strategies accordingly. For example, we can modify the user interface for rating feedback to display one trajectory video at a time with a rating scale. The prompting strategies can be revised to ask users to rate the trajectories with the most distinct features at the initial stage and then proceed to the trajectories with

less familiarity and more disagreements. Future work can further investigate the effectiveness of FARPLS adapting to other feedback types for learning approaches requiring different kinds of human input such as [79].

## 7.3 Limitations and Future Work

Our study has several limitations. First, similar to most existing research on robot-human alignment [14, 62], our study is limited by the representativeness of the participants and data. Although we recruited participants from different backgrounds, including students, researchers, and engineers, we did not have large samples of each group of stakeholders. Also, we did not cover all age groups and all domain expertise groups in robotics or data labeling systems. Future works can compare the performance and experiences of different user groups to gain further insights. Second, because it can be hard to externalize one's internal thinking and reasoning, the criteria and features we summarized from the formative study may not be a complete and fully accurate representation of human values. We may introduce additional sensors [85] or design algorithms [7] to detect implicit preference in human feedback in future work to solve this limitation. Third, due to the limited number of participants and their labeled data, we cannot evaluate the data quality by training and evaluating the reward model. We plan to conduct large-scale crowd-sourcing studies using FARPLS to collect more preference data, evaluate the reward model by training and testing the model on the collected data, and adopt the latest robot task learning algorithms [71] to see the improvement of robot manipulation. Fourth, to ensure the fairness of the between-subjects study, we kept the pool of trajectories for labeling the same for the two conditions and had both groups annotate the entire pool. Hence, we only included 30 trajectories in the pool – a small number compared to many existing robot trajectory datasets – so that the length of the study sessions was reasonable. Also, we did not use any selection strategies to reduce the number of trajectory pairs to be labeled. Future work can incorporate active learning [3], or semi-automatic labeling [24] approaches to reduce the users' workload when more preference data need to be collected over a bigger trajectory pool to train the reward model and evaluate the improvement of the robot task learning.

## 8 CONCLUSION

This paper presents FARPLS, a feature-augmented system for robot trajectory preference labeling. FARPLS is designed to help users establish their criteria and compare the trajectories. Through a formative study, we identified the criteria and features that users care about when labeling the trajectories and the challenges they face when labeling the trajectories. We then derive the design requirements for FARPLS, generate a robot arm dataset, and build a web application with dynamic prompting, adaptive display of features and keyframes, and staged feedback. We conducted a comprehensive user study to evaluate the effectiveness of FARPLS in improving the consistency of human labelers. The results show that FARPLS can significantly improve human labelers' consistency and help labelers in more engaged preference elicitation. The results also show that FARPLS can help users establish the criteria and compare the trajectories with improved user engagement. We also

discuss the generalizability and limitation of FARPLS and provide design considerations for future work in improving the user experience, incorporating algorithmic assistance, and improving the preference dataset for robot learning.

## ACKNOWLEDGMENTS

Many thanks to the anonymous reviewers for their insightful suggestions. We thank all the participants for taking part in the studies and for sharing their valuable feedback. This research was supported in part by the InnoHK funding launched by the Innovation and Technology Commission, Hong Kong SAR and the HKUST & HKPC Joint Laboratory grant HKPC22EG01-A.

## REFERENCES

- [1] Josh Abramson, Arun Ahuja, Federico Carnevale, Petko Georgiev, Alex Goldin, Alden Hung, Jessica Landon, Jirka Lhotka, Timothy Lillicrap, Alistair Muldal, George Powell, Adam Santoro, Guy Sculley, Sanjana Srivastava, Tamara von Glehn, Greg Wayne, Nathaniel Wong, Chen Yan, and Rui Zhu. 2022. Improving Multi-modal Interactive Agents with Reinforcement Learning from Human Feedback. <https://doi.org/10.48550/arXiv.2211.11602> [cs]
- [2] Julie A Adams. 2002. Critical considerations for human-robot interface development. In *Proceedings of 2002 AAAI Fall Symposium*. AAAI Press, North Falmouth, Massachusetts, USA, 1–8.
- [3] Riad Akrou, Marc Schoenauer, and Michèle Sebag. 2012. APRIL: Active Preference Learning-Based Reinforcement Learning. In *Machine Learning and Knowledge Discovery in Databases (Lecture Notes in Computer Science)*, Peter A. Flach, Tijl De Bie, and Nello Cristianini (Eds.). Springer, Berlin, Heidelberg, 116–131. [https://doi.org/10.1007/978-3-642-33486-3\\_8](https://doi.org/10.1007/978-3-642-33486-3_8)
- [4] Pourya Aliasghari, Moojan Ghafurian, Christopher L. Nehaniv, and Kerstin Dautenhahn. 2021. Effects of Gaze and Arm Motion Kinesics on a Humanoid's Perceived Confidence, Eagerness to Learn, and Attention to the Task in a Teaching Scenario. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 197–206. <https://doi.org/10.1145/3434073.3444651>
- [5] Erdem Biyik, Malayandi Palan, Nicholas C. Landolfi, Dylan P. Losey, and Dorsa Sadigh. 2020. Asking Easy Questions: A User-Friendly Approach to Active Reward Learning. In *Proceedings of the Conference on Robot Learning*. PMLR, Virtual, 1177–1190.
- [6] Andreea Bobu, Andi Peng, Pulkit Agrawal, Julie Shah, and Anca D. Dragan. 2023. Aligning Robot and Human Representations. <https://doi.org/10.48550/arxiv.2302.01928> [cs]
- [7] Andreea Bobu, Marius Wiggert, Claire Tomlin, and Anca D. Dragan. 2021. Feature Expansive Reward Learning: Rethinking Human Input. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 216–224. <https://doi.org/10.1145/3434073.3444667>
- [8] Ralph Allan Bradley and Milton E. Terry. 1952. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika* 39, 3/4 (1952), 324–345. <https://doi.org/10.2307/2334029> jstor:2334029
- [9] Daniel Brown, Russell Coleman, Ravi Srinivasan, and Scott Niekum. 2020. Safe imitation learning via fast bayesian reward inference from preferences. In *Proceedings of the 37th International Conference on Machine Learning (2020-11-21)*. PMLR, PMLR, Virtual, 1165–1177. <https://proceedings.mlr.press/v119/brown20a.html>
- [10] Daniel Brown, Wonjoon Goo, Prabhat Nagarajan, and Scott Niekum. 2019. Extrapolating Beyond Suboptimal Demonstrations via Inverse Reinforcement Learning from Observations. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 97)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.). PMLR, Long Beach, California, USA, 783–792. <https://proceedings.mlr.press/v97/brown19a.html>
- [11] Nicholas J. Bryan, Gautham J. Mysore, and Ge Wang. 2014. ISSE: An Interactive Source Separation Editor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 257–266. <https://doi.org/10.1145/2556288.2557253>
- [12] Andrea Bunt, Cristina Conati, and Joanna McGrenere. 2007. Supporting Interface Customization Using a Mixed-Initiative Approach. In *Proceedings of the 12th International Conference on Intelligent User Interfaces*. ACM, Honolulu Hawaii USA, 92–101. <https://doi.org/10.1145/1216295.1216317>
- [13] Ben Carterette, Paul N. Bennett, David Maxwell Chickering, and Susan T. Dumais. 2008. Here or There. In *Advances in Information Retrieval (Lecture Notes in Computer Science)*, Craig Macdonald, Iadh Ounis, Vassilis Plachouras, Ian Ruthven, and Ryen W. White (Eds.). Springer, Berlin, Heidelberg, 16–27. [https://doi.org/10.1007/978-3-540-78646-7\\_5](https://doi.org/10.1007/978-3-540-78646-7_5)
- [14] Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert, Jérémie Scheurer, Javier Rando, Rachel Freedman, Tomasz Korbak, David Lindner, Pedro Freire, Tony Wang, Samuel Marks, Charbel-Raphaël Segerie, Mical Carroll, Andi Peng, Phillip Christoffersen, Mehl Damani, Stewart Slocum, Usman Anwar, Anand Siththanjan, Max Nadeau, Eric J. Michaud, Jacob Pfau, Dmitrii Krasheninnikov, Xin Chen, Lauro Langosco, Peter Hase, Erdem Biyik, Anca Dragan, David Krueger, Dorsa Sadigh, and Dylan Hadfield-Menell. 2023. Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback. <https://doi.org/10.48550/arXiv.2307.15217> arXiv:2307.15217 [cs]
- [15] Kishan Dhananjay Chandan, Jack Albertson, and Shiqi Zhang. 2022. Learning Visualization Policies of Augmented Reality for Human-Robot Collaboration. In *Proceedings of The 6th Conference on Robot Learning*. PMLR, Auckland, New Zealand, 1233–1243.
- [16] Jessie Y. C. Chen, Ellen C. Haas, and Michael J. Barnes. 2007. Human Performance Issues and User Interface Design for Teleoperated Robots. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* 37, 6 (Nov. 2007), 1231–1245. <https://doi.org/10.1109/TSMCC.2007.905819>
- [17] Minsuk Choi, Cheonbok Park, Soyoung Yang, Yonggyu Kim, Jaegul Choo, and Sungsoo Ray Hong. 2019. AILA: Attentive Interactive Labeling Assistant for Document Classification through Attention-Based Deep Neural Networks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300460>
- [18] Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017. Deep Reinforcement Learning from Human Preferences. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc., Long Beach, California, USA. [https://proceedings.neurips.cc/paper\\_files/paper\\_2017/file/d5e2c0ada503c91f91d240d0cd4e49-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper_2017/file/d5e2c0ada503c91f91d240d0cd4e49-Paper.pdf)
- [19] William S Cleveland. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association* 74, 368 (1979), 829–836.
- [20] Vincent Conitzer. 2007. Eliciting Single-Peaked Preferences Using Comparison Queries. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '07)*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/1329125.1329204>
- [21] Daniel Crawl, Jessica Block, Kai Lin, and Ilkay Altintas. 2017. Firemap: A Dynamic Data-Driven Predictive Wildfire Modeling and Visualization Environment. *Procedia Computer Science* 108 (Jan. 2017), 2230–2239. <https://doi.org/10.1016/j.procs.2017.05.174>
- [22] Jingyu Cui, Fang Wen, Rong Xiao, Yuandong Tian, and Xiaou Tang. 2007. EasyAlbum: An Interactive Photo Annotation System Based on Face Clustering and Re-Ranking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. Association for Computing Machinery, New York, NY, USA, 367–376. <https://doi.org/10.1145/1240624.1240684>
- [23] Alexander E Curtis, Tanya A Smith, Bulat A Ziganshin, and John A Elefteriades. 2016. The mystery of the Z-score. *Aorta* 4, 04 (2016), 124–130.
- [24] Daniele De Gregorio, Alessio Tonioni, Gianluca Palli, and Luigi Di Stefano. 2020. Semiautomatic Labeling for Deep Learning in Robotics. *IEEE Transactions on Automation Science and Engineering* 17, 2 (2020), 611–620. <https://doi.org/10.1109/TASE.2019.2938316>
- [25] Michal Dobš, Martin Dekan, Peter Beňo, František Duchoň, and Andrej Babinec. 2022. Evaluation Criteria for Trajectories of Robotic Arms. *Robotics* 11, 1 (2022), 29. <https://doi.org/10.3390/robotics11010029>
- [26] Anca D. Dragan, Kenton C.T. Lee, and Siddhartha S. Srinivasa. 2013. Legibility and Predictability of Robot Motion. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI '13)*. IEEE Press, Tokyo, Japan, 301–308. <https://doi.org/10.1109/HRI.2013.6483603>
- [27] Tim Draws, Alisa Rieger, Oana Inel, Ujwal Gadiraju, and Nava Tintarev. 2021. A Checklist to Combat Cognitive Biases in Crowdsourcing. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing* 9 (2021), 48–59. <https://doi.org/10.1609/hcomp.v9i1.18939>
- [28] Abhishek Dutta and Andrew Zisserman. 2019. The VIA Annotation Software for Images, Audio and Video. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*. Association for Computing Machinery, New York, NY, USA, 2276–2279. <https://doi.org/10.1145/3343031.3350535>
- [29] Samne Elling, Leo Lentz, and Menno de Jong. 2011. Retrospective Think-Aloud Method: Using Eye Movements as an Extra Cue for Participants' Verbalizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1161–1170. <https://doi.org/10.1145/1978942.1979116>
- [30] Teodor Fredriksson, David Issa Mattos, Jan Bosch, and Helena Holmström Olsson. 2020. Data Labeling: An Empirical Investigation into Industrial Challenges and Mitigation Strategies. In *Product-Focused Software Process Improvement (Lecture Notes in Computer Science)*, Maurizio Morisio, Marco Torchiano, and Andreas Jedlitschka (Eds.). Springer International Publishing, Cham, 202–216. [https://doi.org/10.1007/978-3-030-64148-1\\_13](https://doi.org/10.1007/978-3-030-64148-1_13)
- [31] Ujwal Gadiraju, Besnik Fetahu, Ricardo Kawase, Patrick Siehndel, and Stefan Dietze. 2017. Using Worker Self-Assessments for Competence-Based Pre-Selection

- in Crowdsourcing Microtasks. *ACM Transactions on Computer-Human Interaction* 24, 4 (2017), 1–26. <https://doi.org/10.1145/3119930>
- [32] Humberto Garcia Caballero, Michel Westenberg, Binyam Gebre, and Jack van Wijk. 2019. V-Awake: 21st Eurographics/IEEE VGTC Conference on Visualization. *Computer Graphics Forum* 38, 3 (March 2019), 1–12. <https://doi.org/10.1111/cgf.13667>
- [33] Mark E. Glickman and Shane T. Jensen. 2005. Adaptive Paired Comparison Design. *Journal of Statistical Planning and Inference* 127, 1 (Jan. 2005), 279–293. <https://doi.org/10.1016/j.jspi.2003.09.022>
- [34] Abhishek Gupta, Vikash Kumar, Corey Lynch, Sergey Levine, and Karol Hausman. 2020. Relay Policy Learning: Solving Long-Horizon Tasks via Imitation and Reinforcement Learning. In *Proceedings of the Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 100)*, Leslie Pack Kaelbling, Danica Kragic, and Komei Sugiura (Eds.). PMLR, Virtual, 1025–1037. <https://proceedings.mlr.press/v100/gupta20a.html>
- [35] Soheil Habibian, Ananth Jonnaittula, and Dylan P. Losey. 2022. Here's What I've Learned: Asking Questions That Reveal Reward Learning. *ACM Transactions on Human-Robot Interaction* 11, 4 (2022), 40:1–40:28. <https://doi.org/10.1145/3526107>
- [36] Nadia Haddara and Dobromir Rahnev. 2022. The Impact of Feedback on Perceptual Decision-Making and Metacognition: Reduction in Bias but No Change in Sensitivity. *Psychological Science* 33, 2 (2022), 259–275. <https://doi.org/10.1177/09567976211032887>
- [37] J. A. Hartigan and M. A. Wong. 1979. Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28, 1 (1979), 100–108. <https://doi.org/10.2307/2346830> jstor:2346830
- [38] Rachel Holladay, Shervin Javdani, Anca Dragan, and Siddhartha Srinivasa. 2016. Active Comparison Based Learning Incorporating User Uncertainty and Noise.
- [39] Eyke Hüllermeier, Johannes Fürnkranz, Weiwei Cheng, and Klaus Brinker. 2008. Label Ranking by Learning Pairwise Preferences. *Artificial Intelligence* 172, 16–17 (2008), 1897–1916. <https://doi.org/10.1016/j.artint.2008.08.002>
- [40] Eyke Hüllermeier, Johannes Fürnkranz, Weiwei Cheng, and Klaus Brinker. 2008. Label Ranking by Learning Pairwise Preferences. *Artificial Intelligence* 172, 16 (Nov. 2008), 1897–1916. <https://doi.org/10.1016/j.artint.2008.08.002>
- [41] Lise Héroux, Michel Laroch, and K.Lee McGown. 1988. Consumer product label information processing: An experiment involving time pressure and distraction. *Journal of Economic Psychology* 9, 2 (1988), 195–214. [https://doi.org/10.1016/0167-4870\(88\)90051-7](https://doi.org/10.1016/0167-4870(88)90051-7)
- [42] Young-Ho Kim, Diana Chou, Bongshin Lee, Margaret Danilovich, Amanda Lazar, David E. Conroy, Hernisa Kacorri, and Eun Kyung Choe. 2022. MyMove: Facilitating Older Adults to Collect In-Situ Activity Labels on a Smartwatch with Speech. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–21. <https://doi.org/10.1145/3491102.3517457>
- [43] Waldemar W. Koczkodaj. 1998. Testing the Accuracy Enhancement of Pairwise Comparisons by a Monte Carlo Experiment. *Journal of Statistical Planning and Inference* 69, 1 (June 1998), 21–31. [https://doi.org/10.1016/S0378-3758\(97\)00131-6](https://doi.org/10.1016/S0378-3758(97)00131-6)
- [44] Justin Kruger and David Dunning. 1999. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology* 77, 6 (1999), 1121.
- [45] Kostiantyn Kucher, Carita Paradis, Magnus Sahlgren, and Andreas Kerren. 2017. Active Learning and Visual Analytics for Stance Classification with ALVA. *ACM Transactions on Interactive Intelligent Systems* 7, 3 (Oct. 2017), 14:1–14:31. <https://doi.org/10.1145/3132169>
- [46] Caitlin Kuhlman, Diana Doherty, Malika Nurbekova, Goutham Deva, Zarni Phyoo, Paul-Henry Schoenhagen, MaryAnn VanValkenburg, Elke Rundensteiner, and Lane Harrison. 2019. Evaluating Preference Collection Methods for Interactive Ranking Analytics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3290605.3300742>
- [47] Javier Laplaza, Nicolás Rodríguez, J. E. Dominguez-Vidal, Fernando Herrero, Sergi Hernández, Alejandro López, Alberto Sanfelix, and Anaísa Garrell. 2022. IVO Robot: A New Social Robot for Human-Robot Collaboration. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 860–864.
- [48] J. Lee, J. Klie, and I. Gurevych. 2022. Annotation curricula to implicitly train non-expert annotators. *Computational Linguistics* 48 (2022), 343–373. Issue 2. [https://doi.org/10.1162/coli\\_a\\_00436](https://doi.org/10.1162/coli_a_00436)
- [49] Zihao Li, Zhuoran Yang, and Mengdi Wang. 2023. Reinforcement Learning with Human Feedback: Learning Dynamic Choices via Pessimism. [arXiv:2305.18438 \[cs.LG\]](https://arxiv.org/abs/2305.18438)
- [50] Gabrielle Kaili-May Liu. 2023. Perspectives on the Social Impacts of Reinforcement Learning with Human Feedback. arXiv preprint. <https://arxiv.org/abs/2303.02891> arXiv preprint, arXiv:2303.02891
- [51] Evgeni Magid, Roman Lavrenov, and Ilya Afanasyev. 2017. Voronoi-Based Trajectory Optimization for UGV Path Planning. In *2017 International Conference on Mechanical, System and Control Engineering (ICMSC)*. IEEE, St. Petersburg, Russia, 383–387. <https://doi.org/10.1109/ICMSC.2017.7959506>
- [52] Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei, Silvio Savarese, Yuke Zhu, and Roberto Martin-Martín. 2022. What Matters in Learning from Offline Human Demonstrations for Robot Manipulation. In *Proceedings of the 5th Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 164)*, Aleksandra Faust, David Hsu, and Gerhard Neumann (Eds.). PMLR, London, UK, 1678–1690. <https://proceedings.mlr.press/v164/mandlekar22a.html>
- [53] H. B. Mann and D. R. Whitney. 1947. On a Test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics* 18 (1947), 50–60. Issue 1. <https://doi.org/10.1214/aoms/1177730491>
- [54] Yash Dhanpal Mehta, Rohit Ashok Patibanda, and Florian 'Floyd' Mueller. 2018. Arm-A-Dine: Towards Understanding the Design of Playful Embodied Eating Experiences. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 299–313. <https://doi.org/10.1145/3242671.3242710>
- [55] Microsoft. 2021. VoTT. <https://github.com/microsoft/VoTT>
- [56] Daniela Mitterberger, Selen Ercan Jenny, Lauren Vasey, Ena Lloret-Fritsch, Petrus Aeijmelaeus-Lindström, Fabio Gramazio, and Matthias Kohler. 2022. Interactive Robotic Plastering: Augmented Interactive Design and Fabrication for On-Site Robotic Plastering. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–18. <https://doi.org/10.1145/3491102.3501842>
- [57] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. 2018. *Foundations of machine learning*. MIT press, Cambridge, MA.
- [58] Marie Muehlhaus, Marion Koelle, Artin Saberpour, and Jürgen Steimle. 2023. I Need a Third Arm! Eliciting Body-Based Interactions with a Wearable Robotic Arm. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3544548.3581184>
- [59] Jonathan Mumm and Bilge Mutlu. 2011. Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI '11)*. Association for Computing Machinery, New York, NY, USA, 331–338. <https://doi.org/10.1145/1957656.1957786>
- [60] Vivek Myers, Erdem Biyik, Nima Anari, and Dorsa Sadigh. 2022. Learning Multimodal Rewards from Rankings. In *Proceedings of the 5th Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 164)*, Aleksandra Faust, David Hsu, and Gerhard Neumann (Eds.). PMLR, London, UK, 342–352. <https://proceedings.mlr.press/v164/myers22a.html>
- [61] Heramb Nemlekar, Neel Dhanaraj, Angelos Guan, Satyandra K. Gupta, and Stefanos Nikolaidis. 2023. Transfer Learning of Human Preferences for Proactive Robot Assistance in Assembly Tasks. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (Stockholm, Sweden) (HRI '23)*. Association for Computing Machinery, New York, NY, USA, 575–583. <https://doi.org/10.1145/3568162.3576965>
- [62] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., New Orleans, LA, USA [hybrid], 27730–27744. [https://proceedings.neurips.cc/paper\\_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf)
- [63] Li Qian, Jinyang Gao, and H. V. Jagadish. 2015. Learning User Preferences by Adaptive Pairwise Comparison. *Proceedings of the VLDB Endowment* 8, 11 (July 2015), 1322–1333. <https://doi.org/10.14778/2809974.2809992>
- [64] Hannes Ritschel, Andreas Seiderer, Kathrin Janowski, Stefan Wagner, and Elisabeth André. 2019. Adaptive Linguistic Style for an Assistive Robotic Health Companion Based on Explicit Human Feedback. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '19)*. Association for Computing Machinery, New York, NY, USA, 247–255. <https://doi.org/10.1145/3316782.3316791>
- [65] Ork Rooij, Jarke van Wijk, and Marcel Worring. 2010. MediaTable: Interactive Categorization of Multimedia Collections. *IEEE Computer Graphics and Applications* 30, 5 (Sept. 2010), 42–51. <https://doi.org/10.1109/MCG.2010.66>
- [66] Stan Salvador and Philip Chan. 2004. FastDTW: Toward accurate dynamic time warping in linear time and space. In *KDD workshop on mining temporal and sequential data*, Vol. 6. Seattle, Washington, Association for Computing Machinery, New York, NY, USA, 70–80.
- [67] Anara Sandygulova, Aida Amirova, Zhansaule Telisheva, Aida Zhanatkyzy, and Nazerke Rakhymbayeva. 2022. Individual Differences of Children with Autism in Robot-Assisted Autism Therapy. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 43–52.
- [68] Lindsay Sanneman and Julie Shah. 2023. Transparent Value Alignment. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Stockholm Sweden, 557–560. <https://doi.org/10.1145/3568294.3580147>

- [69] Burr Settles. 2011. Closing the Loop: Fast, Interactive Semi-Supervised Annotation with Queries on Features and Instances. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '11)*. Association for Computational Linguistics, USA, 1467–1478.
- [70] R. Sevastjanova, W. Jentner, F. Sperle, R. Kehlbeck, J. Bernard, and M. El-Assady. 2021. Questioncomb: a gamification approach for the visual explanation of linguistic phenomena through interactive labeling. *ACM Transactions on Interactive Intelligent Systems* 11 (2021), 1–38. Issue 3-4. <https://doi.org/10.1145/3429448>
- [71] Daniel Shin, Anca Dragan, and Daniel S. Brown. 2023. Benchmarks and Algorithms for Offline Preference-Based Reward Learning. <https://openreview.net/forum?id=TGuXXlbKsn>
- [72] Bongwon Suh and Benjamin B. Bederson. 2007. Semi-Automatic Photo Annotation Strategies Using Event Based Clustering and Clothing Based Person Recognition. *Interacting with Computers* 19, 4 (July 2007), 524–544. <https://doi.org/10.1016/j.intcom.2007.02.002>
- [73] Daniel Szafir and Danielle Albers Szafir. 2021. Connecting Human-Robot Interaction and Data Visualization. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21)*. Association for Computing Machinery, New York, NY, USA, 281–292. <https://doi.org/10.1145/3434073.3444683>
- [74] Shinichi Tanaka, Young Min Baek, Naohiko Sugita, Takashi Ueta, Yasuhiro Tamaki, and Mamoru Mitsuishi. 2012. Minimum-Jerk Trajectory Generation for Master-Slave Robotic System. In *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, Rome, Italy, 811–816. <https://doi.org/10.1109/BioRob.2012.6290666>
- [75] Almar Van Der Stappen and Mathias Funk. 2021. Towards Guidelines for Designing Human-in-the-Loop Machine Training Interfaces. In *26th International Conference on Intelligent User Interfaces* (College Station TX USA). ACM, College Station, TX, USA, 514–519. <https://doi.org/10.1145/3397481.3450668>
- [76] Chao Wang, Joerg Deignoeller, Pengcheng An, and Julian Eggert. [n. d.]. A User Interface for Sense-making of the Reasoning Process While Interacting with Robots. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany, 2023-04-19). ACM, 1–7. <https://doi.org/10/gss3j7>
- [77] Qianwen Wang, Sehi L'Yi, and Nils Gehlenborg. 2023. DRAVA: Aligning Human Concepts with Machine Learning Latent Dimensions for the Visual Exploration of Small Multiples. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–15. <https://doi.org/10.1145/3544548.3581127>
- [78] Dee Warmath, Dominik Piehlmaier, and Cliff Robb. 2019. The Impact of Shared Financial Decision Making on Overconfidence for Married Adults. *FINANCIAL PLANNING REVIEW* 2, 1 (2019), e1032. <https://doi.org/10.1002/cfp2.1032>
- [79] Nils Wilde, Erdem Biyik, Dorsa Sadigh, and Stephen L. Smith. 2022. Learning Reward Functions from Scale Feedback. In *Proceedings of the 5th Conference on Robot Learning (Proceedings of Machine Learning Research, Vol. 164)*, Aleksandra Faust, David Hsu, and Gerhard Neumann (Eds.). PMLR, London, UK, 353–362. <https://proceedings.mlr.press/v164/wilde22a.html>
- [80] Douglas A Wolfe. 2009. Rank methods. *Wiley Interdisciplinary Reviews: Computational Statistics* 1, 3 (2009), 342–347.
- [81] Bryce Woodworth, Francesco Ferrari, Teofilo E. Zosa, and Laurel D. Riek. 2018. Preference Learning in Assistive Robotics: Observational Repeated Inverse Reinforcement Learning. In *Proceedings of the 3rd Machine Learning for Healthcare Conference (Proceedings of Machine Learning Research, Vol. 85)*, Finale Doshi-Velez, Jim Fackler, Ken Jung, David Kale, Rajesh Ranganath, Byron Wallace, and Jenna Wiens (Eds.). PMLR, Palo Alto, California, USA, 420–439. <https://proceedings.mlr.press/v85/woodworth18a.html>
- [82] Chi-Hsun Wu and Chi-Cheng Jou. 1988. Design of a Controlled Spatial Curve Trajectory for Robot Manipulators. In *Proceedings of the 27th IEEE Conference on Decision and Control*. IEEE, Austin, TX, USA, 161–166 vol.1. <https://doi.org/10.1109/CDC.1988.194289>
- [83] Luyao Yuan, Xiaofeng Gao, Zilong Zheng, Mark Edmonds, Ying Nian Wu, Federico Rossano, Hongjing Lu, Yixin Zhu, and Song-Chun Zhu. 2022. In Situ Bidirectional Human-Robot Value Alignment. *Science Robotics* 7, 68 (July 2022), eabm4183. <https://doi.org/10.1126/scirobotics.abm4183>
- [84] Huixin Zhan, Feng Tao, and Yongcan Cao. 2021. Human-Guided Robot Behavior Learning: A GAN-Assisted Preference-Based Reinforcement Learning Approach. *IEEE Robotics and Automation Letters* 6, 2 (April 2021), 3545–3552. <https://doi.org/10.1109/LRA.2021.3063927>
- [85] Qiping Zhang, Austin Narcomey, Kate Candon, and Marynel Vázquez. 2023. Self-Annotation Methods for Aligning Implicit and Explicit Human Feedback in Human-Robot Interaction. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (New York, NY, USA) (HRI '23)*. Association for Computing Machinery, Stockholm, SE, 398–407. <https://doi.org/10.1145/3568162.3576986>
- [86] Yu Zhang, Bob Coecke, and Min Chen. 2021. MI3: Machine-Initiated Intelligent Interaction for Interactive Classification and Data Reconstruction. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (Sept. 2021), 18:1–18:34. <https://doi.org/10.1145/3412848>
- [87] Yu Zhang, Yun Wang, Haidong Zhang, Bin Zhu, Siming Chen, and Dongmei Zhang. 2022. OneLabeler: A Flexible System for Building Data Labeling Tools. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–22. <https://doi.org/10.1145/3491102.3517612>
- [88] Jing Zhao, Yaxing Duan, Biyun Xie, and Ziqiang Zhang. 2021. FSW Robot System Dimensional Optimization and Trajectory Planning Based on Soft Stiffness Indices. *Journal of Manufacturing Processes* 63 (March 2021), 88–97. <https://doi.org/10.1016/j.jmapro.2020.05.004>
- [89] Banghua Zhu, Jiantao Jiao, and Michael Jordan. 2023. Principled Reinforcement Learning with Human Feedback from Pairwise or \$K\$-wise Comparisons. In *ICLR 2023 Workshop on Mathematical and Empirical Understanding of Foundation Models*. OpenReview.net, Online + Kigali, Rwanda. [https://openreview.net/forum?id=pm\\_WNYd7SP](https://openreview.net/forum?id=pm_WNYd7SP)
- [90] Danny Zhu and Manuela Veloso. 2017. Virtually Adapted Reality and Algorithm Visualization for Autonomous Robots. In *RoboCup 2016: Robot World Cup XX (Lecture Notes in Computer Science)*, Sven Behnke, Raymond Sheh, Sanem Sarie, and Daniel D. Lee (Eds.). Springer International Publishing, Cham, 452–464. [https://doi.org/10.1007/978-3-319-68792-6\\_38](https://doi.org/10.1007/978-3-319-68792-6_38)
- [91] Yuke Zhu, Josiah Wong, Ajay Mandlekar, and Roberto Martín-Martín. 2020. Robosuite: A Modular Simulation Framework and Benchmark for Robot Learning. [arXiv:2009.12293](https://arxiv.org/abs/2009.12293)

## A PICKPLACECANS FEATURE DEFINITIONS

Table 9 shows the formula for each feature as a time series and as a scalar in a trajectory complementary to Table 2. We stack all feature time series for each criterion to form a criterion vector series:

$$\text{safety}_i(t) = [ \quad \text{num\_collision}_i(t), \text{dis\_to\_left}_i(t), \\ \text{dis\_to\_right}_i(t), \text{dis\_to\_front}_i(t), \\ \text{dis\_to\_back}_i(t), \text{dis\_to\_table}_i(t), \\ \text{contact\_force}_i(t) \quad ];$$

$$\text{efficiency}_i(t) = [ \quad \text{speed}_i(t), \text{eef\_posi}(t), \\ \text{can\_posi}(t), \text{pseudo\_cost}_i(t) \quad ];$$

$$\text{task\_quality}_i(t) = [ \quad \text{speed\_smoothness}_i(t), \\ \text{trajectory\_smoothness}_i(t), \\ \text{orientation}_i(t), \text{grasp\_position}_i(t) \quad ];$$

which represent three criteria, *Safety*, *Efficiency*, and *Task Quality*, for trajectory  $i$  at  $t$ -th step.

**Table 9: Formula definitions for each feature as a time series and as a scalar in a trajectory. In the formulas,  $i$  represents the trajectory,  $s_i$  denotes the total number of steps (i.e., the number of states) in trajectory  $i$  and  $t \in \{0, 1, \dots, s_i\}$  denotes the step-index in the trajectory series.**

Feature	Formula	
Safety		
Collision	Time Series	$num\_collisions_i(t)$
	Scalar	$max\_num\_collisions_i = \max\{collision_i(t)\}_{t=0}^{s_i}$
Distance (to table edges)	Time Series	$dis\_to\_left_i(t), dis\_to\_right_i(t), dis\_to\_front_i(t),$ and $dis\_to\_back_i(t).$
	Scalar	$min\_dis\_to\_edge_i = \min\{dis\_to\_left_i(t), dis\_to\_right_i(t), dis\_to\_front_i(t), dis\_to\_back_i(t)\}_{t=0}^{s_i}$
Distance (to table surface)	Time Series	$dis\_to\_table_i(t)$
	Scalar	$max\_height\_to\_table_i = \max\{dis\_to\_table_i(t)\}_{t=0}^{s_i}$
Contact force	Time Series	$eef\_force_i(t)$
	Scalar	$max\_eef\_force_i = \max\{eef\_force_i(t)\}_{t=0}^{s_i}$
Efficiency		
Speed	Time Series	$speed_i(t)$
	Scalar	$avg\_speed_i = \frac{1}{s_i} \sum_{t=0}^{s_i} speed_i(t)$
	Time Series (1)	$eef\_pos_i(t)$
	Time Series (2)	$can\_pos_i(t)$
Path Length	Scalar (1)	$reach\_length_i = \sum_{t=1}^{s_1}  eef\_pos_i(t) - eef\_pos_i(t-1) $
	Scalar (2)	$grasp\_length_i = \sum_{t=s_1+1}^{s_2}  eef\_pos_i(t) - eef\_pos_i(t-1) $
	Scalar (3)	$grasp\_length_i = \sum_{t=s_1+1}^{s_3}  eef\_pos_i(t) - eef\_pos_i(t-1) $
	Scalar	$total\_time_i(t) = \frac{s_i}{fps}$
Time	Keyframe (1)	The step $s_1$ that the end-effector gripped the can when picking up. (e.g., Figure 3e)
	Keyframe (2)	The step $s_2$ that the end-effector releases the can when placing. (e.g., Figure 3f)
Power Usage	Time Series	$pseudo\_cost_i(t) = \sum_{\tau=0}^t \sum_{j=1}^{dof}  q_j(\tau) $
	Scalar	$pseudo\_cost_i = \sum_{\tau=0}^{s_i} \sum_{j=1}^{dof}  q_j(\tau) $
Task Quality		
Speed	Time Series	$speed\_smoothness_i(t) = \frac{1}{t} \sum_{\tau=0}^t \sqrt{\mathbf{a}(\tau)^2}$
	Scalar	$speed\_smoothness_i = \frac{1}{s_i} \sum_{\tau=0}^{s_i} \sqrt{\mathbf{a}(\tau)^2}$
Trajectory Smoothness	Time Series	$trajectory\_smoothness_i(t) = \frac{1}{t} \sum_{\tau=0}^t \arccos \frac{\mathbf{x}(\tau) \cdot \mathbf{x}(\tau+1)}{ \mathbf{x}(\tau)   \mathbf{x}(\tau+1) }$
	Scalar	$trajectory\_smoothness_i(t) = \frac{1}{s_i} \sum_{\tau=0}^{s_i} \arccos \frac{\mathbf{x}(\tau) \cdot \mathbf{x}(\tau+1)}{ \mathbf{x}(\tau)   \mathbf{x}(\tau+1) }$
Orientation	Time Series	$orientation_i(t) = eef\_ori\_mat_i(t)^{-1} * can\_ori\_mat_i(t)$
Grasp Position	Time Series	$grasp\_position_i(t) = can\_pos_i(t) - eef\_pos_i(t)$