# How Can Everyday Users Efficiently Teach Robots by Demonstration?

MARAM SAKR, University of British Columbia, Canada
ZHIKAI ZHANG, University of British Columbia, Canada
BENJAMIN LI, University of British Columbia, Canada
HAOMIAO ZHANG, University of British Columbia, Canada
H.F. MACHIEL VAN DER LOOS, University of British Columbia, Canada
DANA KULIĆ, Monash University, Australia
ELIZABETH CROFT, University of Victoria, Canada

Learning from Demonstration (LfD) is a framework that allows lay users to easily program robots. However, the efficiency of robot learning and the robot's ability to generalize to task variations hinge upon the quality and quantity of the provided demonstrations. Our objective is to guide human teachers to provide more effective demonstrations, thus facilitating efficient robot learning. To achieve this, we propose to use a measure of uncertainty, namely task-related *information entropy*, as a criterion for suggesting informative demonstration examples to human teachers to improve their teaching skills. This approach seeks to minimize the requisite number of demonstrations by enhancing their distribution throughout the workspace. In a conducted experiment ($N = 24$), an augmented reality (AR)-based guidance system was employed to train novice users to produce additional demonstrations from areas with the highest entropy within the workspace. These novice users were trained for a few trials to teach the robot a generalizable task using a limited number of demonstrations. Subsequently, the users' performance after training was assessed first on the same task (retention) and then on a new task (transfer) without guidance. The results indicate a substantial improvement in robot learning efficiency from the teacher's demonstrations, with an improvement of up to 198% observed on the novel task. Furthermore, the proposed approach was compared to a state-of-the-art heuristic rule and found to improve robot learning efficiency by 210% compared to the heuristic rule. The scripts used in this paper are available on GitHub.

## 1 INTRODUCTION

In Learning from Demonstration (LfD), robots acquire skills through examples provided by human demonstrators, offering a promising avenue for robot learning from everyday users. LfD allows users to convey task information naturally, contrasting with direct task programming [2, 5]. Such natural interactions are particularly advantageous for novice users, who may lack the technical expertise required for programming through conventional interfaces or a deep understanding of the underlying systems [18].

While LfD builds upon the successes of standard Machine Learning (ML) methods, which have been applied successfully across a diverse array of applications [18], teaching robots through human demonstrations introduces unique challenges. These include constraints such as limited human patience and the generation of low-quality and inefficient data [4]. The size and diversity of the training data set will determine the speed and accuracy of learning, including its generalization characteristics. Moreover, it has been shown that data modifications

Authors' addresses: Maram Sakr, University of British Columbia, Canada; Zhikai Zhang, University of British Columbia, Canada; Benjamin Li, University of British Columbia, Canada; Haomiao Zhang, University of British Columbia, Canada; H.F. Machiel Van der Loos, University of British Columbia, Canada; Dana Kulić, Monash University, Australia; Elizabeth Croft, University of Victoria, Canada.


ACM 2573-9522/2025/5-ART
https://doi.org/10.1145/3737892

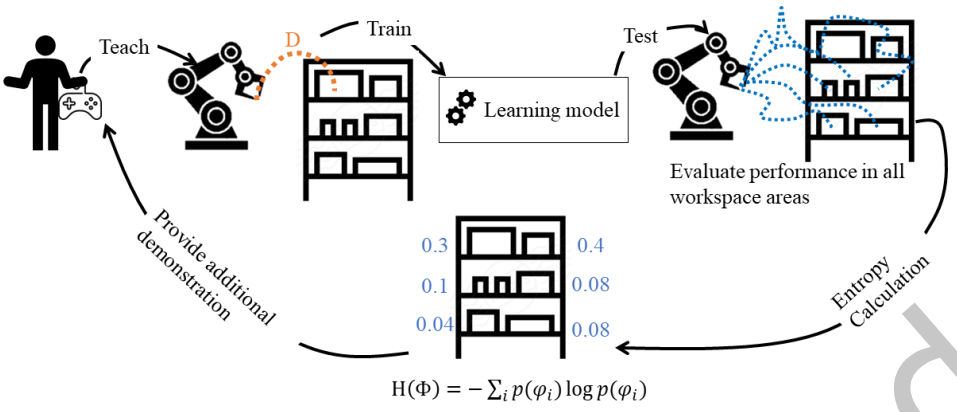



Fig. 1. **System Overview:** The system uses information entropy to guide users to the areas of highest uncertainty in Learning from Demonstration (LfD). The model is initially trained with a single demonstration, then calculates the uncertainty across the task space. The area with the highest uncertainty is suggested to the user for additional demonstrations. This process repeats until the robot achieves certainty in executing the task across the entire task space.

often impact generalization more profoundly than modifications to the learning algorithm itself [22]. Thus, this paper focuses on optimizing the input (i.e., demonstration data) to the learning algorithm, rather than altering the algorithm itself.

Human teaching strategies are typically optimized for human learners and may not align perfectly with the needs of machine learners [13]. Nevertheless, human teachers can adjust to a specific learner's requirements [30]. To facilitate this adaptation, we formulate the problem as a machine teaching problem, aiming to identify the smallest set of demonstrations necessary to achieve effective learning and generalization. To this end, we introduce the concept of "teaching guidance", which provides human teachers with targeted instructions to influence their selection of examples, focusing on those that are most informative for the specific learner (e.g., a robot).

We aim to help novice teachers provide demonstrations efficiently by using information entropy as the criterion for selecting demonstrations for robot learning and generalization through the following contributions: i) introducing task-dependent information entropy into the LfD pipeline to define uncertainty across all task space areas, as shown in Fig. 1; ii) proposing the integration of Augmented Reality (AR) with the entropy-based guidance system as a training framework to help novice users better identify informative demonstrations; iii) evaluating the proposed training framework in a user study; and iv) comparing robot learning efficiency using our information entropy approach with a state-of-the-art heuristic rule [46].

## 2 RELATED WORK

Teaching robots through demonstration poses challenges for lay users as they may struggle to provide demonstrations that are both useful [30, 44] and sufficient [19, 46]. Therefore, it is important to guide these users in providing informative demonstrations that enable the robot to learn and generalize beyond the examples provided. Generalization of the demonstrated trajectories is one of the most important challenges in LfD systems. The provided demonstrations are merely samples from the entire task space and are used to train a learning model to capture the underlying *latent* space, rather than simply replicating them. To achieve this generalization, it is crucial to consider both the diversity [23] and the distribution [42] of the provided demonstrations over the task space. In other words, demonstrations should be as diverse as possible to maximize coverage of the task space while minimizing the number of demonstrations needed (i.e., achieving efficiency). While some researchers

have taken an algorithmic perspective to improve overall learning efficiency in reinforcement learning [38, 53], our main focus in this work is to propose a human-centered approach aimed at improving users' teaching skills through training and guidance, thereby enhancing robot learning efficiency.

## 2.1 Teaching Guidance and Users Training

While significant research has been conducted on policy-learning methods for LfD, much less attention has been paid to the quality of teaching and strategies to improve it [5, 16]. Effective teaching, even when instructing humans, is inherently challenging as it requires understanding the learner's knowledge, employing suitable teaching strategies, and adjusting based on feedback. Teaching robots presents an even greater challenge due to the complexity of understanding and addressing the robot's learning requirements.

Cakmak and Takayama [11] explored the use of instructional materials to guide novice teachers in teaching robots through kinesthetic teaching and spoken dialogue interfaces. Ajaykumar et al. [1] proposed a curriculum-based training program for novice users, focusing on teaching users how to manipulate robots. However, both the instructional materials and the curriculum did not address the efficiency or quantity of demonstrations. Mohseni-Kabir et al. [37] designed an interactive system for teaching hierarchical task structures, where participants were given explicit instructions for teaching each task. While effective in controlled scenarios, such approaches are impractical for scaling robotics to everyday use, as it is infeasible for experts to teach every end user how to program robots for every desired task. Instead, our approach focuses on how users teach tasks to robots without explicit guidance from roboticists. We emphasize enabling users to optimally select a set of demonstrations that allows robots to learn and generalize tasks effectively.

From a human social learning perspective, two critical observations about teaching guidance were highlighted by Chernova and Thomaz [18]: (i) effective teachers maintain a mental model of learner understanding and (ii) learners assist teachers by expressing their internal state through communicative acts, both verbal and non-verbal. While these principles are central to human teaching, they are often overlooked in LfD systems, leaving significant opportunities for improvement. For example, in an LfD user study, participants expressed a need for transparency in the learner's knowledge to better understand the robot's learning progress [51]. Similarly, Huang et al. [28] demonstrated how robots providing informative demonstrations to humans can enhance the human's understanding of robot objectives in autonomous driving tasks.

These insights inform the design of our training and guidance framework for lay users. Our proposed system helps users understand the robot's learning progress and the impact of demonstration quantity and distribution on the robot's performance. By doing so, the framework aims to minimize the time and effort required from users while ensuring they provide a minimal yet effective set of demonstrations to achieve robust learning and generalization.

## 2.2 Active Learning

Robots may become more proactive in their learning by requesting additional demonstrations [48] or seeking clarification [12]. This is what is called "active learning" which makes robot learning adaptive by allowing the robot to request guidance when facing uncertainties or novel situations. Unlike passive LfD, this approach is valuable in cases with limited initial demonstrations, helping the robot learn correct actions in real-time and reducing unsafe or suboptimal decisions [49]. However, this approach places a burden on the user to be available for further queries, which may be impractical.

Efforts to reduce this burden have included minimizing human involvement, as demonstrated in [41, 54, 57], who propose optimizing human time, effort and attention as a cost function. At each step, the robot must decide whether to query the user or rely on its autonomous controller. However, these systems depend on the availability of an autonomous controller to replace human input in certain scenarios, which may be impractical or

infeasible for some tasks. Dass et al. [21] proposed automating parts of the data collection and asking for human help only in ambiguous scenarios. Others tried to make the queries user-friendly to make it easier for users to respond [6]. Trinh et al. [55] introduced a self-assessment mechanism, where robots evaluate the sufficiency of the demonstrations they receive rather than relying on human feedback. DAgger, introduced by Ross et al. [42], combats the distributional shift and unseen state generalization issues prevalent in behavioral cloning by supplementing demonstrations with expert-labelled actions in newly encountered states, thus improving robustness. Subsequent work, such as that by Zhang and Cho [56], builds on DAgger with methods to reduce the number of expert labels required during training. Similarly, Chen et al. [17] propose an approach to streamline reinforcement learning (RL) by enabling the RL agent to actively query for demonstrations during training. Other methods, such as those by Dai et al. [20] and Tao et al. [52], leverage expert demonstrations and curriculum learning to enhance demonstration efficiency in RL. All these strategies assume access to expert demonstration data or that the demonstrator is an expert, aiming to reduce the need for expert intervention during training. In contrast, this paper proposes a framework for training non-expert users to provide effective demonstrations, allowing robots to learn efficiently from users without prior expertise. Our framework operates without any expert in the loop or reliance on expert demonstrations.

Others have proposed incremental learning approaches that enable robots to learn and refine their skills from human demonstrations without requiring expert demonstrations or interventions [29, 50]. Mehta et al. [36] introduced a framework, StROL, to ensure the stability and robustness of robots learning online from human input, addressing challenges such as variability in feedback and task instability. Calinon and Billard [15] introduced an incremental approach in which users teach robots gestures by demonstrating tasks recorded with a motion tracking system. The robot then performs the task, and users refine robot's skills through kinesthetic teaching. This approach relies on the teacher to determine the location of the next demonstration after observing the robot's performance in various locations. However, this can be particularly challenging for novice teachers, who may struggle to identify the locations for robot evaluation and decide where to provide subsequent demonstrations.

To address this challenge, Sena and Howard [46] proposed heuristic rules for determining demonstration locations in the task space. These rules include: (i) starting with one demonstration from any location in the task space, (ii) continuing to provide demonstrations within 4 cm of the first demonstration until it is surrounded by successful test points, and (iii) adding further demonstrations within 4 cm of successful test points in areas with the highest number of failed test points. While these rules are effective for their specific experimental setup, they are difficult to generalize to other tasks. To overcome these limitations and create a more general framework, we propose using uncertainty measure (information entropy) to guide users in providing efficient demonstrations. This approach aims to simplify the teaching process for novice users, making it more efficient and adaptable across different tasks.

## 2.3 Machine Teaching

A closely related field of research that considers optimal teaching of machine learning systems is *machine teaching*, which aims to design the optimal training data to drive the learning algorithm to a target model [58]. Two prominent methods within this field are the "teaching dimension model", which represents the minimum number of instances a teacher must reveal to uniquely identify any target concept [25], and the "curriculum learning principle", which presents training examples in a structured sequence, progressing from simple to complex concepts [3]. These approaches open avenues for exploring optimal teaching strategies specifically for robots. For instance, Khan et al. [30] compared the teaching dimension model and curriculum learning principle in teaching a simple 1D classification problem, while Cakmak and Lopes [10] studied optimal teaching in sequential decision tasks. Building on this foundation, Brown and Niekum [9] reformulated inverse reinforcement learning

(IRL) as a machine teaching problem, aiming to identify the smallest set of demonstrations required to effectively convey the reward structure of a task.

A critical component of machine teaching is selecting the most informative data points to maximize learning efficiency. In robotics, this is analogous to identifying the most impactful demonstrations in the task space. Selecting these demonstrations is conceptually similar to choosing exploration areas in reinforcement learning (RL) [33] or defining the query that will maximize the information gained about the user's preference in active preference learning [7]. Maximizing information entropy is a popular approach that has been used in both reinforcement learning and preference learning [34]. Typically, information entropy is used to characterize the performance of an agent through the definition of a reward function, where the next action is chosen based on the observed entropy maxima. Bıyık et al. [6] utilize the information entropy to select the queries that achieve a balance between information gain and the ability for the user to answer the question confidently. Haarnoja et al. [26] proposed a soft actor-critic policy that aims to maximize expected reward while also maximizing entropy. That is, to succeed at the task while acting as randomly as possible to increase information gain. In robot learning from demonstration problems, the efficiency of user-provided demonstrations is a key factor in determining how quickly and effectively the task will be learned by the robotic agent.

Building on these ideas, we formulate the problem of selecting the most efficient set of demonstrations for a robot to learn from as a machine teaching problem. By leveraging information entropy, we aim to minimize teaching costs and alleviate the teaching burden on users, especially novices. Furthermore, we integrate the proposed entropy-based approach into an augmented reality (AR) guidance system. This system directs users to areas of high uncertainty in the task space, ensuring that additional demonstrations provide maximum benefit for the robot's learning and generalization.

## 3 OBJECTIVES

The objective of this research is to enhance human teaching skills to provide efficient demonstrations to robots. Therefore, we propose using information entropy as the criterion for suggesting to the human demonstrator the most informative demonstrations from which the robot can learn. The research questions we aim to address are as follows:

(1) Does information entropy improve robot learning efficiency by guiding demonstration selection?
(2) Does an entropy-based guidance system enhance users' performance in teaching robots through demonstration?
(3) How much visual information should the guidance system provide to improve the user's teaching efficiency?

To address the first question, we compared an entropy-based guidance system with a heuristic guidance system as proposed in [46] (described in Section 2). To tackle the second question, we incorporated the proposed entropy-based guidance system into a training framework for novice users to assess whether their teaching skills improved—specifically in terms of the number and distribution of provided demonstrations—after training. For the third question, we conducted a user study comparing two visual guidance systems: 1) single point suggestion and 2) heatmap representation of the task space's entropy values. This comparison aims to determine how the amount of visual information provided affects the users' teaching efficiency.

## 4 PROPOSED APPROACH

In the field of LfD, determining the most effective demonstration set for a robot is a critical challenge that directly impacts learning efficiency, generalization, and adaptability to new tasks. This problem can be formulated as a *machine teaching problem*, where the objective is to identify the optimal training set that enables a robot to approximate a target policy with the fewest possible demonstrations. To address this, we propose using

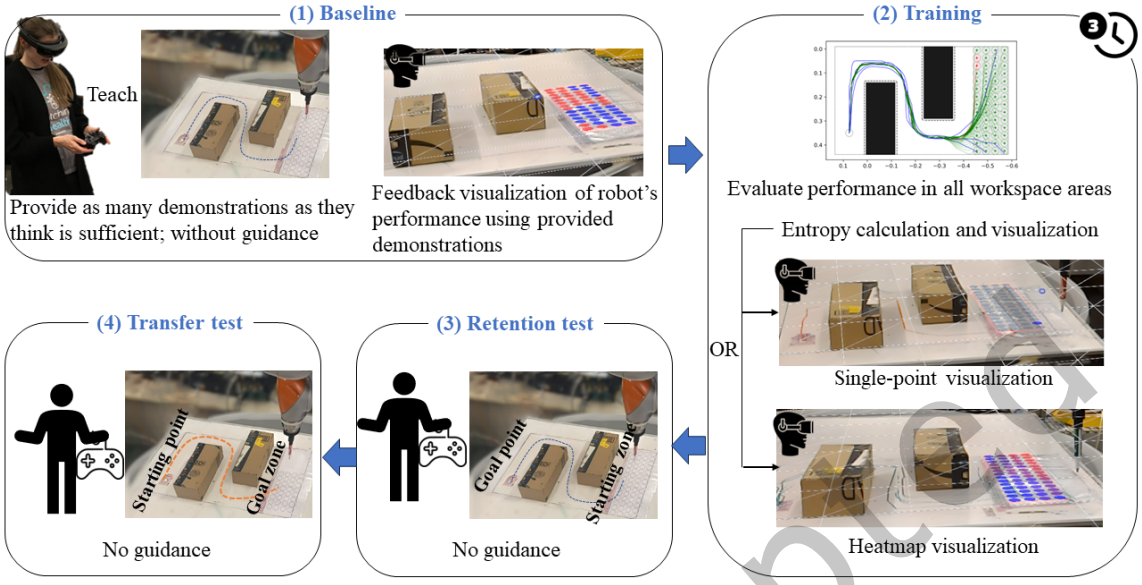



Fig. 2. Overview of the user study procedure. **(1) Baseline:** The user provides as many demonstrations as they think would be sufficient for the robot to learn the task. **(2) Training:** The user is guided to the uncertain areas for three trials. **(3) Retention test:** The user is evaluated on the same task without guidance. **(4) Transfer test:** The user teaches the robot to move from a specific starting point to a wider goal zone without guidance.

uncertainty measures to guide the selection of the most informative demonstrations, enabling efficient and scalable robot learning.

### 4.1 Formulating the Machine Teaching Problem

Following the machine teaching framework [59], our objective is to find a demonstration set $\mathcal{D}$ that allows the learner (the robot) to approximate a target policy $\theta^*$ as closely as possible. The machine teaching problem can be framed as a bilevel optimization problem as follows:

$$\min_{\mathcal{D},\hat{\theta}} |\hat{\theta} - \theta^*|^2 + \eta|\mathcal{D}|, \tag{1}$$

where $|\hat{\theta} - \theta^*|^2$ quantifies the divergence between the learner's policy $\hat{\theta}$ and the target policy $\theta^*$, and $\eta$ is a weighting parameter that balances the trade-off between the teaching cost $|\mathcal{D}|$ (the cardinality of the demonstration set) and learning accuracy. The learner's optimization task, defined as the inner problem, is to minimize the empirical loss:

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} \sum_{(x,y) \in \mathcal{D}} \mathcal{L}(x, y, \theta) + \lambda|\theta|^2, \tag{2}$$

where $\mathcal{L}(x, y, \theta)$ represents the empirical loss between the true outcome $y$ and the model output $\theta(x)$, and $\lambda$ is the regularization parameter penalizing model complexity.

The outer optimization in equation 1 is the teacher's problem. The teacher aims to bring the learner model $\hat{\theta}$ close to the target model $\theta^*$ while also using a small teaching set ($|\mathcal{D}|$). The inner optimization described by equation 2 is the learner's machine learning problem [59]. The machine teaching problem can be reformulated as follows:

$$\min_{\mathcal{D},\hat{\theta}} \quad TeachingRisk(\hat{\theta}) + \eta\ TeachingCost(\mathcal{D}) \tag{3}$$

$$\text{s.t.} \quad \hat{\theta} = MachineLearning(\mathcal{D}) \tag{4}$$

where $TeachingRisk(\hat{\theta})$ is defined to quantify the teacher's dissatisfaction with the learner's current model. The target model $\theta^*$ may be incorporated into the teaching risk function, but the formulation does not explicitly require a target parameter $\theta^*$. In this paper, we adapt the concept of teaching risk to focus on the learner's generalizability across a predefined task space. Additionally, the *TeachingCost* is defined as the cardinality of the provided demonstrations as in [9]. By simultaneously minimizing both teaching cost and teaching risk, our approach optimizes the efficiency of the robot's learning process.

### 4.2 Incorporating Entropy to Optimize Demonstrations

To optimize the teaching cost, we propose using information entropy [47] as a measure of uncertainty to guide the demonstration selection process. Entropy enables the identification of regions where the robot is uncertain about its ability to achieve success. These high-uncertainty regions are prioritized for additional demonstrations, reducing the total number of demonstrations required while improving generalization. Mathematically, entropy is defined by the Shannon formula as follows:

$$H(\Phi) = -\sum_{i=1}^{n} p(\varphi_i) \log p(\varphi_i) \tag{5}$$

where $p(\varphi_i)$ is the estimated probability of a specific outcome $\varphi_i$ from a set of possible outcomes $\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$. High entropy values (e.g., when $p(\varphi) \approx 0.5$) indicate maximum uncertainty, making these regions prime candidates for additional demonstrations.

Uncertainty in robot learning arises from multiple sources, including task model dynamics, variability in learning algorithms, and inconsistencies in robot sensing and control. Each of these sources contributes to the robot's overall uncertainty in achieving task success. In the proposed approach, information entropy is used to quantify this aggregated uncertainty and identify high-uncertainty regions where additional demonstrations can improve learning efficiency. The probability $p(\varphi_i)$, representing the likelihood of each outcome, can be estimated using empirical success rates, model-based methods like regression or Gaussian Processes, or heuristic-based scoring derived from task-specific features.

### 4.3 Application to Task-Specific Success Criteria

In this paper, we address uncertainty arising from task dynamics and learning algorithms by defining task-specific success criteria and evaluating performance across multiple rollouts for stochastic learning policies or a single rollout for deterministic policies. This approach is applicable to any task with well-defined success criteria across the workspace. Here, we focus on a task where a robot learns to navigate a constrained workspace, a scenario common in domestic, industrial, and exploration settings. Such tasks typically involve reaching a goal location while avoiding obstacles and remaining within a defined workspace. Success for these tasks is defined by the following criteria:

(1) The robot's end effector reaches the target location.

(2) No collisions occur with obstacles in the environment.
(3) The robot remains within the defined workspace boundary throughout the trajectory.

The proposed entropy-based approach is designed to be broadly applicable across tasks where success criteria can be explicitly defined over a workspace. In this paper, we demonstrate the method using a robot navigation task in a constrained 2D workspace—a setting common in domestic, industrial, and exploratory applications. While the current evaluation focuses on a simple goal-reaching task, the formulation itself is not limited to this domain. By defining appropriate task-specific success features, the same approach can be extended to more complex tasks such as object manipulation, multi-step planning, or dynamic environments.

In this framework, uncertainty is addressed through measurable criteria derived from the robot's ability to complete the task successfully. These criteria are used to estimate the probability of success in each region of the task space and compute entropy as a measure of uncertainty. Although the learning algorithm used in our implementation is a deterministic model (TP-GMM), the entropy-based strategy is agnostic to the underlying policy representation and can be applied with other LfD methods, including behavioral cloning or reinforcement learning-based imitation methods.

For our navigation task, we define success based on the following interpretable and generalizable criteria:

Fig. 1 illustrates an example task using the proposed entropy-based approach. In this example, a user demonstrates to the robot how to retrieve items from a shelf. The objective is to minimize the number of demonstrations while achieving effective generalization across the shelf area. After an initial demonstration, the robot learns a model of the task, and the workspace is discretized into small regions. The robot's performance is then evaluated in each region based on the success criteria defined above. Given the binary nature of the generated trajectory (success or failure) in each region, entropy for each region is calculated as follows:

$$H(r) = -p(r)\log p(r) - (1 - p(r))\log(1 - p(r)) \tag{6}$$

where $p(r)$ denotes the estimated probability of success within region $r$, while $1 - p(r)$denotes the estimated probability of failure within region $r$. This probability of success $p(r)$ for each region is calculated as a weighted sum of task-specific features, averaged over the number of rollouts:

$$p(r) = \frac{\sum_{n=1}^{N} \mathbf{w}^T \mathbf{x}}{N_{runs}}, \quad \sum_i \mathbf{w}_i = 1, \tag{7}$$

where $\mathbf{w}$ represents the weights assigned to each task-specific feature $\mathbf{x}$, and $N_{runs}$ denotes the total number of rollouts. The definition of success features and desired trajectory characteristics is task-dependent. The weight vector allows prioritization of certain features based on their relative importance to the task. For the navigation task in this study, the estimated probability of success is computed using the success criteria defined above as follows:

$$p(r) = \frac{\sum_{n=1}^{N} (w_1 \cdot (1 - d_{goal\_norm}) + w_2 \cdot n_{ncollision\_norm} + w_3 \cdot n_{inside\_norm})}{N_{runs}}, \tag{8}$$

where $d_{goal\_norm}$ is the normalized distance from the trajectory endpoint to the goal, $n_{ncollision\_norm}$ is the normalized number of trajectory points that are not in collision with obstacles, and $n_{inside\_norm}$ is the normalized number of trajectory points that are within the defined workspace boundaries. $N_{runs}$ is the number of rollouts. Each feature is normalized by its minimum and maximum observed value across all trajectories to ensure consistent scaling. The weights $w_1, w_2$, and $w_3$ are experimentally set to $w_1 = 0.3$, $w_2 = 0.4$, and $w_3 = 0.3$, reflecting a comparable level of significance of all task success features for this specific task.

To ensure efficient learning, regions with the highest entropy are prioritized for additional demonstrations. The next region to receive a demonstration is selected by identifying the region with the maximum entropy value:

$$r_{\text{next}} = \arg\max_{r \in \text{Regions}} H(r). \tag{9}$$

This iterative process continues until the entropy across all regions is sufficiently low, signifying the robot's confidence in consistently performing the task throughout the workspace. Notably, the proposed entropy-based approach is independent of the user's expertise level. If initial demonstrations are of low quality, the system can prompt users to provide additional demonstrations to support effective learning and robust generalization. Moreover, the entropy metric is computed independently of the underlying learning algorithm and task complexity.

### 4.4 Visual Guidance Using Augmented Reality (AR)

To guide users to areas of high uncertainty, we utilize a Microsoft HoloLens AR headset [31]. The headset visualizes entropy values within the task space, giving users intuitive feedback on the robot's learning progress. Providing the user with more insight into the learning process may lead to more effective teaching. However, providing too much information may overwhelm and hinder the user, especially one that is untrained. To find an adequate amount of information to be presented in the visual interface about the robot's learning progress, we proposed two visualization schemes (Fig. 2-2) as follows:

- **Heatmap Visualization**: Displays entropy values across the task space as a color-coded heatmap.
- **Single-Point Visualization**: Highlights only the most uncertain point where additional demonstration is needed.

Using AR for guidance makes the system applicable to both 2D and 3D tasks, enhancing situational awareness by overlaying feedback directly on the task, unlike the computer screen-based feedback in [46].

### 4.5 Task Learning

For task learning, we used the Task Parameterized Gaussian Mixture Model (TP-GMM) [14]. TP-GMMs have been extensively used for LfD [39, 40], and they show good generalization using a limited set of demonstrations. TP-GMMs model a task with parameters that are defined by a sequence of coordinate frames. In a $D$ dimensional space, each task parameter/coordinate frame is given by an $A \in \mathbb{R}^{D \times D}$ matrix indicating its orientation and a $b \in \mathbb{R}^{D}$ vector indicating its origin, relative to the global frame. A $K$-component mixture model is fitted to the data in each local frame of reference. Each GMM is described by $\left(\pi_k, \mu_k^{(j)}, {}_k^{(j)}\right)$, referring to the prior probabilities, mean, and covariance matrices for each component $k$ in frame $j$, respectively.

To use the local models for trajectory generation, they must be projected back into the global frame of reference and then combined into one global model. Continuous trajectories can then be generated from the global mixture model using Gaussian Mixture Regression (GMR); Calinon [14] provides further details. A Bayesian Information Criterion (BIC) was used to define the optimal number of $K$-Gaussian components to fit the demonstrations.

In our experimental task, we used the same hyperparameters for the TP-GMM model to ensure deterministic outcomes. This enabled us to evaluate the model with a single rollout, simplifying the computational process. Since the system was evaluated with users in an online setting, minimizing computational time was essential to maintain responsiveness and usability.

## 5 EVALUATION

To evaluate the proposed entropy system, a preliminary experiment in simulation comparing the proposed approach and one of the state-of-the-art heuristic guidance rules is presented. It aims to study the impact of the proposed approach on robot learning efficiency. After that, a user study is presented to explore the contribution of the proposed approach to improving human teaching skills.

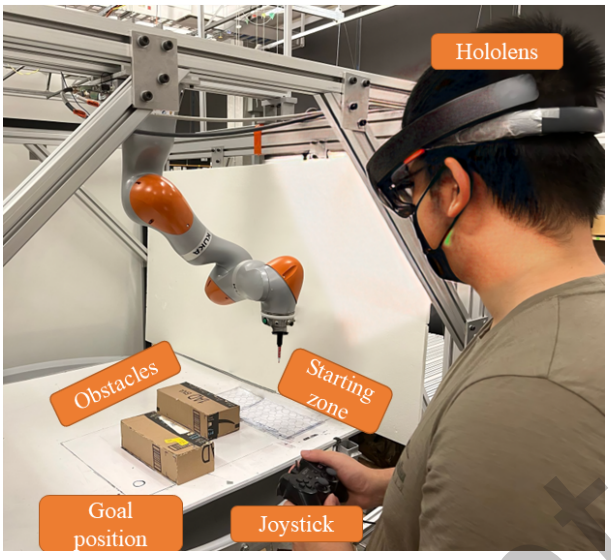



Fig. 3. Experimental setup used in the user study.

## 5.1 Experimental Task

Fig. 3 represents the experimental task, in which a Kuka IIWA LBR14 robot is taught to navigate a 2D maze by drawing a trajectory on the whiteboard from a starting zone to a target point. Similar success criteria as in Section 4 are used. The robot should avoid collisions with the obstacles in the workspace and it should not go outside the workspace specified by a two-dimensional bounding rectangle (45 cm by 72 cm) drawn on the whiteboard. The starting zone is defined by a 45 cm by 15 cm rectangle. In the experiment, the starting zone is discretized to a 4 x 14 test grid with 56 circles separated by 0.5 cm. This task was designed to resemble a general robot navigation scenario in a constrained workspace. Its complexity was intentionally balanced to ensure it was neither too simple—requiring no training for users to grasp how to teach it to the robot—nor too challenging for the targeted novice users.

Participants use the joystick to teach the robot to navigate through the maze. The teaching process is finished if the robot can generate a "successful" trajectory for 90% of the points in the starting zone to the goal position. The success criteria of the trajectories are similar to what is defined in the example task in Section 4. Microsoft Hololens is used to overlay a geometrically accurate starting zone on the real environment, offering user feedback on the robot's progress and guiding efficient demonstration, as shown in Fig. 2.

A modified version of this task was used in the transfer test for evaluating the *user's* performance in new tasks. The training task was flipped; the goal point became a starting point and the starting zone became the goal zone. This goal zone was shortened to a 4 x 8 test grid, as shown in Fig. 2.

## 5.2 Performance Metrics

### *5.2.1 Quantitative measures.*

(1) **Task Completion Time:** We measured the total time, $t$, required by the user to complete the experiment task.
(2) **Number of Demonstrations:** We measured the number of demonstrations $|\mathcal{D}|$ that achieve at least 90% coverage of the starting zone.

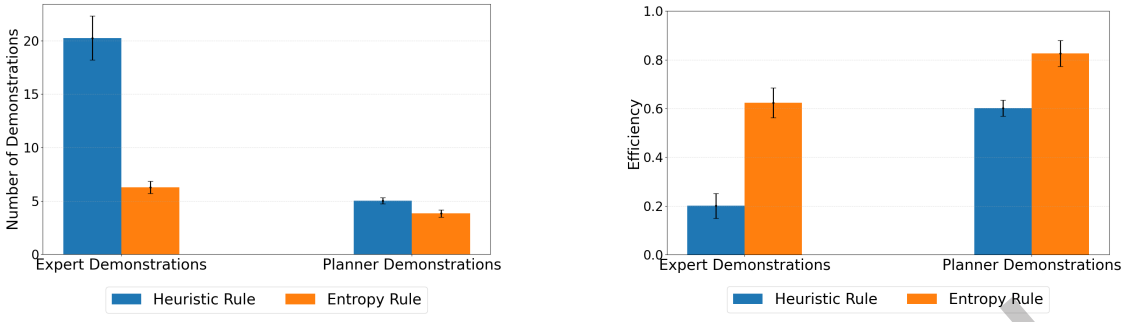



Fig. 4. The average number of demonstrations and robot learning efficiency for both the entropy rule and the heuristic rule with planner and expert demonstrations.

(3) **Teaching Efficacy:** Efficacy refers to the ability to perform a task to a satisfactory or expected degree. Here, we used the teaching efficacy defined in [46], which is the ratio between the successful points,$|\hat{I}|$, and the total number of tested points, $|I|$, ($\epsilon = \frac{|\hat{I}|}{|I|}$).

(4) **Teaching Efficiency:** Efficiency in any given application is often context-dependent; however, typically it is desirable to minimize the total number of demonstrations required because this is correlated with both the time spent teaching and the space needed to store data. Here, we used the teaching efficiency defined in [46], which is the efficacy normalized by the number of the provided demonstrations, $|\mathcal{D}|$,

$$\eta = \frac{\epsilon}{|\mathcal{D}|}. \tag{10}$$

*5.2.2 Qualitative measures.*

(1) **NASA Task Load Index (NASA-TLX):** This questionnaire [27] is composed of six questions asking the participants to rate their perceived task load in six different aspects on a 21-point scale.

(2) **System Usability Scale (SUS).** This scale evaluates the usability of the proposed training system. The SUS [8] questionnaire includes ten questions asking the participant to rate different aspects of the system's usability on a 10-point scale. An overall score is then calculated.

## 5.3 Preliminary Experiment: Guidance Rule in Simulation

In this experiment our goal is to compare the entropy-based guidance rule with the heuristic guidance rule proposed in [46] (described in Section 2). Various factors influence robot learning and generalization, including the quality, quantity, and distribution of provided demonstrations. Here, we specifically focus on the guidance rule that governs the "number" and "distribution" of provided demonstrations, while keeping all other factors constant. To ensure consistent demonstration quality, we employed the motion planner RRT-Connect [32] to generate demonstrations for both guidance rules. Trajectories were generated from all points within the starting zone to the goal point and saved for the experiment. Additionally, to simulate a real-case scenario involving human demonstrations, we collected expert demonstrations, performed by one of the paper's authors, for all points within the starting zone of the experimental task shown in Fig. 3.

For the comparison study, for each guidance rule, we developed a Python script to guide the robot's learning process. The experimental setup was simulated with RViz, the standard tool for visualization and interaction with robot applications implemented in ROS. The script initially proposes a point in the starting zone, and its corresponding saved trajectory (either from the motion planner or expert demonstrations) is retrieved. Then, the TP-GMM model is trained by this trajectory and tested with all points in the starting zone to identify the

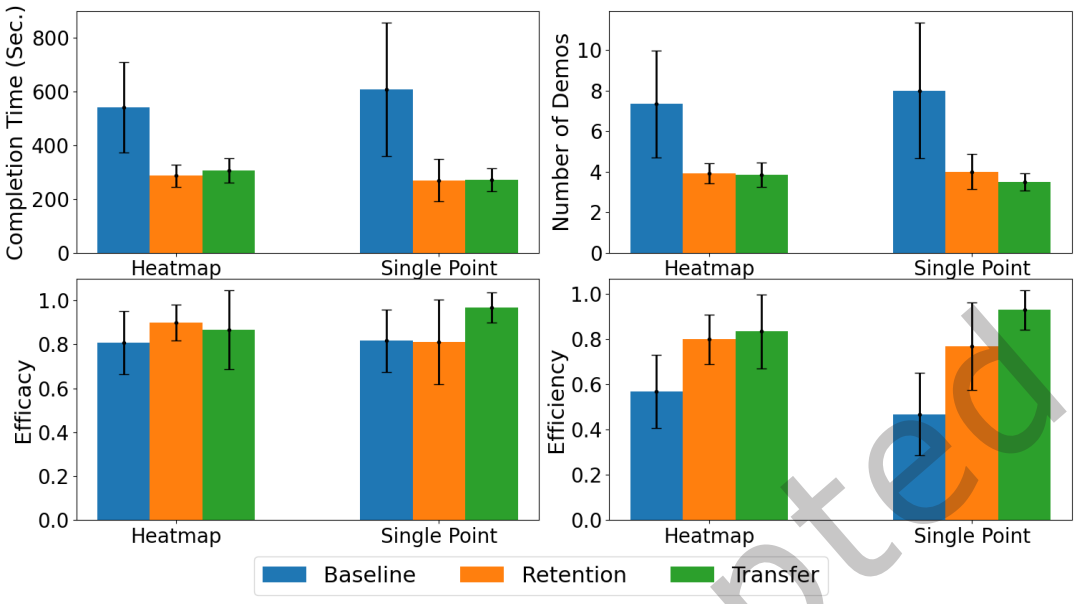



Fig. 5. Performance metrics for comparing heatmap and single-point groups in baseline, retention, and transfer tests.

successful and failed trajectories. Using the guidance rule, the script selects the next point in the starting zone for the next demonstration. These steps are repeated until the learning model generates successful trajectories for 90% or more of the points in the starting zone. The same procedure is repeated 56 times for both guidance rules until every point in the task space has been tested as the initial point. Finally, the two guidance rules are compared with respect to the number of demonstrations needed to generate successful trajectories for 90% or more of the starting zone.

#### 5.3.1 *Hypotheses.*

The experimental hypotheses are chosen to test whether robot learning efficiency, Equation 10, improves using the proposed approach compared to the heuristic rule. The alternative hypothesis is formally defined as:

- **H1(a)**: For robot learning using planner demonstrations, learning efficiency using the entropy guidance rule is significantly higher than learning efficiency using the heuristic rule.
- **H1(b)**: For robot learning using human expert-generated demonstrations, learning efficiency using the entropy guidance rule is significantly higher than learning efficiency using the heuristic rule.

#### 5.3.2 *Results.*

Fig. 4 shows the mean and 95% confidence interval of the number of demonstrations and robot learning efficiency in 56 trials using the heuristic rule and entropy rule with both expert and planner demonstrations. Overall, using the entropy rule achieves a higher robot learning efficiency than the heuristic rule, for both planner and expert demonstrations. A 2x2 Bayesian ANOVA was conducted to investigate the impact of a) guidance rule, and b) demonstration type on robot learning efficiency. Both the guidance rule ($BF_{10} = 7.43 \times 10^{15}$) and demonstration type ($BF_{10} = 2.82 \times 10^{13}$) show a large effect on robot learning efficiency. In addition, robot learning efficiency using either guidance rule strongly depends on the demonstration type (the quality of the provided demonstrations) ($BF_{10} = 184.11$). These results support **H1(a)** and **H1(b)**.

Interestingly, a slight decrease in the demonstration's quality (i.e., from the planner to the expert demonstrations) causes a drastic decrease in the robot learning efficiency by 200% using the heuristic rule, while using the proposed

entropy rule the efficiency decreased by only 33.9%. This shows the importance of the guidance rule in robot learning from demonstrations and its contribution to the robot learning efficiency which cannot be compensated only by improving the demonstration quality. Thus, it is important to guide the users in providing an efficient set of demonstrations from which the robot can learn.

## 5.4 User Study Experiment

From the simulation experiment results, we found that the entropy guidance rule significantly improves robot learning efficiency. Therefore, we utilized this rule in a training framework for users to guide them to provide an efficient set of demonstrations for the robot. Fig. 2 shows an overview of the user study procedure as follows:

(1) **Baseline:** The user provides as many demonstrations as they think would be sufficient for the robot to generate successful trajectories from any point in the starting zone to the goal position. The provided demonstrations are used to train a TP-GMM model and evaluate the performance which is provided back to the user via the Hololens as visualized feedback. The blue circles represent successfully learnt trajectories while the red circles represent failed learnt trajectories. The baseline phase represents the initial interaction of the users with the robot.
(2) **Training:** The user provides an initial demonstration, and then the learning model is evaluated across the starting zone. The entropy is calculated in each area and visualized through the HoloLens as a heatmap or single-point, according to the training group. The user provides an additional demonstration in the highest entropy point and the entropy visualization is updated, and so on. This continues until the robot learns successful trajectories for 90% or more of the starting zone. Each user was trained for three trials.
(3) **Retention Test:** This step is similar to the baseline to evaluate the performance after training.
(4) **Transfer Test:** The user teaches the robot to move from a specific starting point to a goal zone without guidance, described in Section 5.1.

The users were asked to fill in the NASA-TLX questionnaire [27] after their baseline and retention tests to see the effect of the training on their perceived task load. They were also asked to fill in the SUS questionnaire [8] after the training to get their subjective evaluation of the proposed training system. Lastly, a semi-structured interview was conducted before and after training to inquire about users' reasoning behind their selection of the number of provided demonstrations and their respective placements. These interviews were conducted with the intention of delving into users' mental models of robot learning before and after training.

### *5.4.1 **Hypotheses**.*

The experimental hypotheses are chosen to test whether the user's teaching skills measured through robot learning efficiency are significantly improved using the proposed approach. They are defined as:

- **H2**: Robot learning efficiency will be significantly improved after training the human teacher using the proposed approach (over both groups).
- **H3**: The heatmap visualization group will achieve a significantly higher improvement in robot learning efficiency than the single-point visualization group on the retention test (i.e., same training task).
- **H4**: The heatmap visualization group will achieve a robot learning efficiency that is significantly higher than the single-point visualization group on the transfer test (i.e., on a novel task).

The user study is a between-participants design, so each participant is assigned to one group: either the heatmap group or the single-point group. A priori power analysis was conducted using G*Power version 3.1.9.7 [24] to determine the minimum sample size required to test the study hypotheses. Results indicated the required sample size to achieve 80% power for detecting a medium effect size (Cohen's $f = 0.3$), and a type I error rate $\alpha = 0.05$, is $N = 24$, or 12 participants per condition. Accordingly, the results reported below are for a population of 24 participants (14 male, 10 female; ages $\mu = 24, \sigma = 4$). We recruited participants for our user study through advertisements posted on the University of British Columbia (UBC) campus and social media. Prior to conducting

the study, we obtained research ethics approval from UBC's Behavioural Research Ethics Board (application ID H20-03740-A001). We obtained informed consent from each participant before commencing the experiment.

*5.4.2* ***Results.***

**Quantitative Measures**

Fig. 5 depicts the mean and 95% confidence interval for all performance metrics across baseline, retention, and transfer tests for both the heatmap and single-point groups. Overall, there is a noticeable improvement in all performance metrics during the retention and transfer tests when compared to the baseline performance. Efficiency improved by 140% and 164% from the baseline to the retention for both heatmap and single-point groups, respectively, and by 146% and 198% from the baseline to the transfer test for both heatmap and single-point groups, respectively. Interestingly, users' performance improved more in the transfer test than in the retention test compared to their baseline performance. Given that the efficiency metric encapsulates both the number of demonstrations and efficacy, the statistical analysis was centred on efficiency.

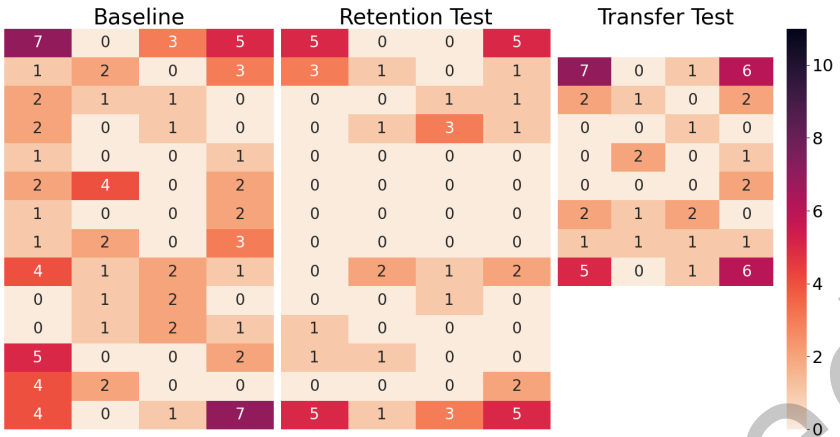
Bayesian mixed analysis of variance (ANOVA) showed a strong impact of the training on efficiency when comparing the baseline to the retention test ($BF_{10} = 1387.16$); supporting **H2**. The same analysis showed inconclusive results on the impact of the visualization type on efficiency ($BF_{10} = 0.4$); no support for **H3** and **H4**. Furthermore, the Bayesian independent samples t-test showed inconclusive results on the comparison between the transfer test efficiency in both heatmap and single point groups ($BF_{10} = 0.60$). This result highlights the effectiveness of the proposed entropy-based guidance system, irrespective of how the entropy information is visually presented to the user.

We also investigated the distribution of the demonstrations over the task space across the baseline, retention and transfer tests. This shows the different strategies users adopted to distribute their demonstrations over the task space. Fig. 6 shows a heatmap representing the number and locations of the provided demonstrations across different stages of the experiment for both heatmap and single-point groups. The number of the provided demonstrations notably decreased from the baseline to the two tests. The median number of baseline demonstrations was 6 and 7, respectively, for the heatmap and single-point groups, decreasing to a median of 4 in the retention and transfer tests for both groups. Compared to the baseline which has a higher number of demonstrations from points in the interior of the start zone, the retention and transfer test demonstrations are concentrated at the corners. This highlights the change in users' strategy for selecting the number and locations of demonstrations.

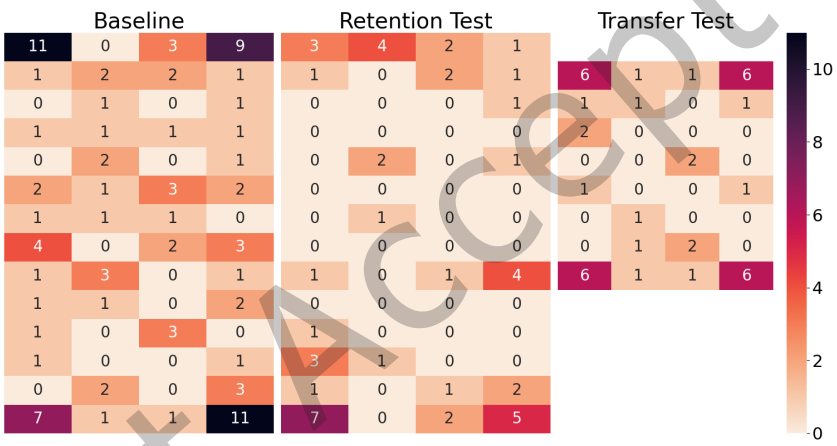
The proposed approach implicitly addresses the issue of low-quality demonstrations from novice users. If a user provides low-quality demonstrations, the system will request additional demonstrations to improve task learning and generalization. For instance, Fig. 7 illustrates the pattern of robot learning efficacy for each provided demonstration from three different users along with the distribution of the demonstrations over the starting zone. The first user provided high-quality demonstrations, so only three demonstrations were sufficient for the learning model to generalize across the entire task space. The second user provided mixed-quality demonstrations, hence a higher number of demonstrations was required to achieve generalization across the task space, with an efficacy of 96%. The third user provided low-quality demonstrations, thus the learning model only generalized across 43% of the task space with six demonstrations. Here, we used task learning performance as a measure of quality, as proposed in our previous paper [44].

**Qualitative Measures**

**Perceived Task Load:** Fig. 8 illustrates the box plots for both the heatmap and single-point groups following the baseline and retention tests. It shows that both effort and frustration have decreased after training in the heatmap group while only the effort decreased in the single-point group. In contrast to the quantitative measures, users' perceived performance after training was lower than before training in both groups. In general, the task load metrics statistically indicate no significant difference before and after the training. This suggests that the training did not have an adverse impact on users' perceived task load or lead to overwhelming experiences.




(a) Heatmap group

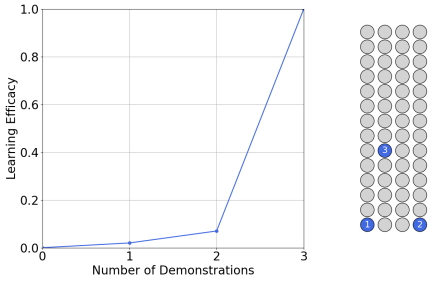



(b) Single-point group

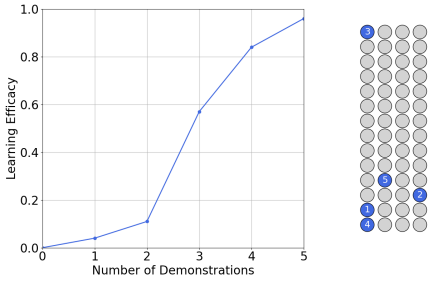
Fig. 6. The number and distribution of the demonstrations in the starting zone for the baseline, retention, and transfer tests for both the (a) heatmap group and (b) single-point group. Users provided a high number of demonstrations and distributed them across the starting zone in the baseline. After training, they provided a lower number of demonstrations focused on the corners.



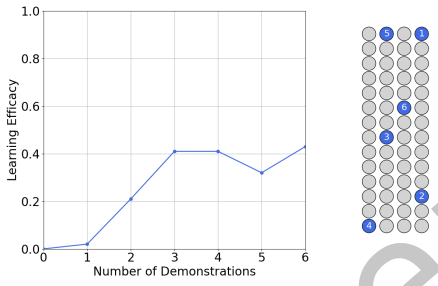
**System Usability:** The usability score for the heatmap group is ($M = 75.21; 95\%CI, 64.99 to 85.42$), while in the single point group ($M = 62.71; 95\%CI, 55.89 to 69.53$). According to the global benchmark for SUS created by Sauro and Lewis through surveying 446 studies spanning different types of systems, the mean given score is $68 \pm 12.5$ [45]. When comparing this global benchmark mean score with the scores obtained from the two training groups, the observed differences were inconclusive (Bayesian one-sample t-test $BF_{10} = 0.75, BF_{10} = 0.89$, respectively). This suggests that both training approaches are usable from the user's perspective.




(a) High-quality demonstrations example




(b) Mixed-quality demonstrations example




(c) Low-quality demonstrations example

Fig. 7. An example of high-, mixed-, and low-quality demonstrations, illustrating their relationship to learning efficacy and their distribution across the task space.



**Interviews:** Participants employed diverse strategies when determining the number and placements of demonstrations during the baseline phase. Among the 24 participants, 19 opted to establish a grid of points, strategically encompassing both the corners and intermediary positions. As articulated by one participant: *"Well, overall, I just wanted to get the corners and then a few different points along the diagonals in between. I just felt like it kind of covers the majority of the grid"*, which is reflected in Fig. 6. In contrast, two other participants opted for a completely random selection of demonstrations, favoring a higher count of demonstrations (18 and 23) to ensure comprehensive coverage of the starting area. Another participant thought that providing a dense set of points would allow the robot to learn the distance between the points in the starting zone and generalize well to other areas. Another participant provided three demonstrations that formed a triangle. They thought that the two points forming the base of the triangle would specify the width for the learning algorithm to interpolate in, while the third point would specify the height for the algorithm.

After training, the participants approached the task differently and efficiently thought about the number and distribution of the provided demonstrations. Almost all participants provided an average number of demonstrations in retention equal to the average number of demonstrations they provided in the training, regardless of the training condition. As one participant said, *"After the second training trial, I realized that I don't need a really large number of points as long as I pick the four corners of the rectangular area"*. In addition, there was an agreement from several participants in the heatmap group that visualizing the generalization area of each point in the training was very helpful in deciding the location of the next demonstration in the retention test. Some

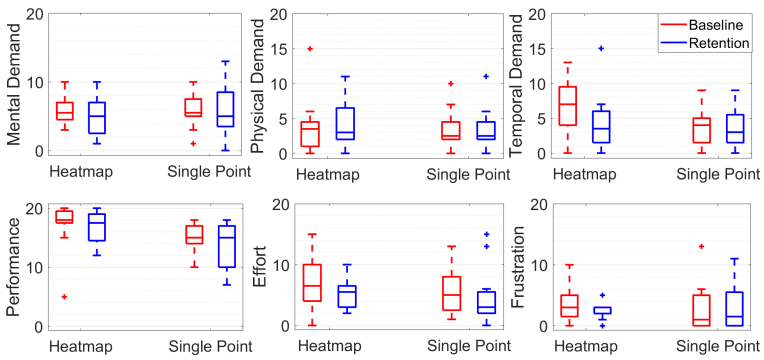



Fig. 8. Participants response to NASA-TLX [27] questions after baseline and retention test. There is no significant difference before and after the training.

participants counted the number of generalized points around one demonstration in the training and tried to distance the demonstrations in the retention by that number of points.

## 6 DISCUSSION

This study demonstrated that a few trials of interactive training and guidance for lay users significantly improved their teaching skills, which in turn enhanced robot learning and generalization efficiency. Notably, this learning and generalization occurred online, using demonstrations from teachers with no prior knowledge of robotics or machine learning algorithms. By using the proposed training framework, users develop an understanding of what demonstrations the robot requires to learn efficiently, without necessarily understanding how the learning process occurs. Thus, the overarching goal of this work is realized by decreasing the teaching cost through guiding users to provide a small number of demonstrations that achieve wider generalization.

From the preliminary experiment results and statistical analysis, there is support for **H1(a)** and **H1(b)**. Robot learning efficiency using the information entropy rule was significantly higher than using the heuristic rule with both expert demonstrations and motion planner demonstrations. Furthermore, using the same task as in [46] but with a larger scale, we found a failure case for the heuristic rule: it happens in 61% of the total experiment trials. This failure case happens whenever all the surrounding points are demonstrated and the learning algorithm performs poorly in learning and generalization. This failure case is expected to be even worse with novices who tend to provide low-quality demonstrations [2] that are negatively affecting robot learning and generalization [44].

The proposed entropy-based approach has the potential for generalization to a wide range of tasks where success criteria can be defined across a task workspace. In complex scenarios such as object manipulation, multi-agent coordination, or dynamic obstacle avoidance, entropy can quantify uncertainty arising from task-specific factors like grasp stability, coordination reliability, or environmental dynamics. By defining measurable success criteria for each task, entropy identifies high-uncertainty regions where additional demonstrations are most informative. For instance, in object manipulation, entropy can prioritize demonstrations in regions where stable grasps are uncertain due to variations in object geometry or surface friction. Similarly, in dynamic obstacle avoidance, entropy can guide attention to areas with high variability in obstacle motion patterns. This flexibility makes the entropy-based approach a valuable tool for optimizing demonstration selection in diverse and complex robotic tasks, reducing teaching costs while improving generalization. While the entropy-based strategy is

theoretically applicable to any LfD algorithm that produces success/failure outcomes, further empirical validation is needed to confirm its robustness across different learning models and task domains.

Furthermore, to ensure scalability in more complex tasks, an exhaustive evaluation of all possible starting regions is not necessary. Efficient heuristic-based or sampling techniques can be employed to identify representative high-uncertainty regions, significantly reducing computational overhead. By concentrating demonstrations in these targeted areas, the proposed approach achieves scalability without compromising its generalization benefits, even across larger or more intricate task spaces.

The user study design effectively captures users' initial performance with the robot before any training. By providing feedback on their performance, the need for training is emphasized. During the training phase, users receive feedback on the robot's generalization performance after each demonstration. This helps them understand how the distribution of their demonstrations influences the robot's learning. Evaluating the robot's overall performance across the entire task space helps avoid the biases that can arise from assessing only a subset of tasks, as noted in [46]. Because each user completes three training trials, they may become more familiar with the robot's learning process and the way its policy evolves. As a result, it can be difficult to fully separate this learning effect from the impact of the proposed guidance framework without additional experiments or analysis. In future work, incorporating control groups or exploring different feedback modalities could help disentangle these effects and provide a clearer understanding of the guidance framework's contribution to improved teaching performance.

From the user study results and statistical analysis, there is support for **H2** but no support for **H3** or **H4**. Both training groups showed a significant improvement in learning efficiency from the baseline to the retention and transfer tests. This underscores the importance of training novice users before teaching the robot, consistent with prior work [11, 16]. Additionally, this study shows the benefit of the proposed entropy guidance system regardless of the visualization used in training. Notably, both training groups achieved similar performance in retention and transfer tests. This is because most participants had a preconception of creating a grid of points to cover the starting zone to help the robot learn better, as shown in Fig. 6. Subsequently, post-training, participants recognized the feasibility of achieving high efficiency with a reduced number of demonstrations, contrary to their initial assumptions. As shown in Fig. 5, the efficacy before and after training is similar. This finding contrasts with previous work [46], which found that most participants tend to underestimate the number of demonstrations required to teach the robot effectively.

Regarding the usability of the guidance system, the heatmap group reported a higher SUS score than the single-point group, while both training groups showed an inconclusive difference with the global mean score. We posit that this is because participants in the heatmap condition had access to a larger amount of information during training than the single-point group. The heatmap visualization provided feedback on robot learning progress and an efficient way for deciding their next demonstration. Only one participant commented that they wished to have a highlight of the highest uncertainty point in the heatmap. This is because they sometimes found the colors in the heatmap were very close to each other, making it hard to define the highest uncertainty point. Participants in the single-point group did not have access to the robot's learning progress and therefore had to depend on the demonstration point suggestions to find a pattern on how to efficiently teach the robot.

A surprising observation emerged concerning users' perceived performance post-training. Despite a significant quantitative improvement in performance, 8 out of the 24 participants reported a decline in perceived performance compared to before training. This discrepancy was investigated by analyzing their interview data related to their approach to determining the number and locations of demonstrations. These participants experienced significant shifts in their strategies after training, which impacted their confidence and perception of performance. For instance, one participant provided 16 demonstrations in the baseline. After training, they provided four demonstrations in the retention test and remarked in the interview, *"I just used four points, although I'm not sure I covered 90%, I don't know."* Others changed their strategies from dense point sets to a more distributed

approach, albeit with skepticism. These findings suggest the potential need for adaptive training tailored to individual participants' requirements, especially for those with strong pre-training strategy convictions, aligning with previous research [44].

Another interesting observation is that one participant without prior knowledge of robotics realized the importance of the demonstration's quality in conjunction with the number and location of the provided demonstrations for robot learning and generalization [35]. They tried different starting points in the training and then realized that providing smooth demonstrations without abrupt corners allows the robot to learn and generalize faster with a smaller number of demonstrations, as shown in Fig. 7. However, it is challenging to extrapolate that all participants would naturally arrive at this realization without proper training, as in [43]. This underscores the necessity for an instructional framework to guide participants in offering demonstrations of both high quality and optimal efficiency, echoing the emphasis of previous research [2].

## 7 CONCLUSION AND FUTURE WORK

Toward our goal of achieving efficient robot learning and generalization, we proposed using information entropy as a criterion for guiding human teachers to provide the next demonstration that achieves the highest efficiency. This proposed approach reduces the teaching dimension by suggesting the minimum number of demonstrations that achieve the highest learning efficiency. The approach was validated in two experiments to explore its contribution to both efficient robot learning and generalization, as well as improving novices' teaching skills. We found that the proposed approach significantly improves robot learning efficiency compared to the state-of-the-art heuristic rule. Moreover, information entropy was integrated into two training schemes and tested in a user study. The results show a significant improvement in robot learning efficiency after training in both groups.

We believe that the results of this paper open up several directions for future research. The significant improvement in efficiency by merely guiding users to strategically distribute the demonstrations suggests that guiding users to provide high-quality demonstrations, as in [43], along with their good distribution, could further boost learning efficiency. It would also be interesting to test the proposed approach in facilities with real users without controlling the conditions. For instance, users would have the freedom to decide how long they need guidance to ensure they provide the most beneficial demonstrations to the robot.



## REFERENCES


[1] Gopika Ajaykumar, Gregory D Hager, and Chien-Ming Huang. 2023. Curricula for teaching end-users to kinesthetically program collaborative robots. *Plos one* 18, 12 (2023), e0294786.

[2] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and autonomous systems* 57, 5 (2009), 469–483.

[3] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*. 41–48.

[4] Aude G Billard, Sylvain Calinon, and Rüdiger Dillmann. 2016. Learning from humans. *Springer handbook of robotics* (2016), 1995–2014.

[5] Erik A Billing and Thomas Hellström. 2010. A formalism for learning from demonstration. *Paladyn, Journal of Behavioral Robotics* 1, 1 (2010), 1–13.

[6] Erdem Bıyık, Malayandi Palan, Nicholas C Landolfi, Dylan P Losey, and Dorsa Sadigh. 2019. Asking easy questions: A user-friendly approach to active reward learning. *arXiv preprint arXiv:1910.04365* (2019).

[7] Erdem Biyik and Dorsa Sadigh. 2018. Batch active preference-based learning of reward functions. In *Conference on robot learning*. PMLR, 519–528.

[8] Julian Brooke. 2006. SUS - A quick and dirty usability scale.

[9] Daniel S Brown and Scott Niekum. 2019. Machine teaching for inverse reinforcement learning: Algorithms and applications. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7749–7758.

[10] Maya Cakmak and Manuel Lopes. 2012. Algorithmic and human teaching of sequential decision tasks. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

[11] Maya Cakmak and Leila Takayama. 2014. Teaching people how to teach robots: The effect of instructional materials and dialog design. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*. 431–438.

[12] Maya Cakmak and Andrea L Thomaz. 2012. Designing robot learners that ask good questions. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 17–24.
[13] Maya Cakmak and Andrea L Thomaz. 2014. Eliciting good teaching from humans for machine learners. *Artificial Intelligence* 217 (2014), 198–215.
[14] Sylvain Calinon. 2016. A tutorial on task-parameterized movement learning and retrieval. *Intelligent service robotics* 9, 1 (2016), 1–29.
[15] Sylvain Calinon and Aude Billard. 2007. Incremental learning of gestures by imitation in a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*. 255–262.
[16] Sylvain Calinon and Aude G Billard. 2007. What is the teacher's role in robot programming by demonstration?: Toward benchmarks for improved learning. *Interaction Studies* 8, 3 (2007), 441–464.
[17] Si-An Chen, Voot Tangkaratt, Hsuan-Tien Lin, and Masashi Sugiyama. 2020. Active deep Q-learning with demonstration. *Machine Learning* 109, 9 (2020), 1699–1725.
[18] Sonia Chernova and Andrea L Thomaz. 2014. Robot learning from human teachers. *Synthesis lectures on artificial intelligence and machine learning* 8, 3 (2014), 1–121.
[19] Sonia Chernova and Manuela Veloso. 2009. Interactive policy learning through confidence-based autonomy. *Journal of Artificial Intelligence Research* 34 (2009), 1–25.
[20] Siyu Dai, Andreas Hofmann, and Brian Williams. 2021. Automatic curricula via expert demonstrations. *arXiv preprint arXiv:2106.09159* (2021).
[21] Shivin Dass, Karl Pertsch, Hejia Zhang, Youngwoon Lee, Joseph J Lim, and Stefanos Nikolaidis. 2022. Pato: Policy assisted teleoperation for scalable robot data collection. *arXiv preprint arXiv:2212.04708* (2022).
[22] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.
[23] Benjamin Eysenbach, Abhishek Gupta, Julian Ibarz, and Sergey Levine. 2019. Diversity is All You Need: Learning Skills without a Reward Function. In *International Conference on Learning Representations*. https://openreview.net/forum?id=SJx63jRqFm
[24] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.
[25] Sally A Goldman and Michael J Kearns. 1995. On the complexity of teaching. *J. Comput. System Sci.* 50, 1 (1995), 20–31.
[26] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer Dy and Andreas Krause (Eds.). PMLR, 1861–1870. https://proceedings.mlr.press/v80/haarnoja18b.html
[27] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
[28] Sandy H Huang, David Held, Pieter Abbeel, and Anca D Dragan. 2019. Enabling robots to communicate their objectives. *Autonomous Robots* 43, 2 (2019), 309–326.
[29] Michael Kelly, Chelsea Sidrane, Katherine Driggs-Campbell, and Mykel J Kochenderfer. 2019. Hg-dagger: Interactive imitation learning with human experts. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 8077–8083.
[30] Faisal Khan, Bilge Mutlu, and Jerry Zhu. 2011. How do humans teach: On curriculum learning and teaching dimension. *Advances in neural information processing systems* 24 (2011).
[31] Bernard C Kress and William J Cummings. 2017. Towards the Ultimate Mixed Reality Experience: HoloLens Display Architecture Choices. In *SID Symposium Digest of Technical Papers*, Vol. 48. 127–131.
[32] James J Kuffner and Steven M LaValle. 2000. RRT-connect: An efficient approach to single-query path planning. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, Vol. 2. IEEE, 995–1001.
[33] Chunmao Li, Xuanguang Wei, Yinliang Zhao, and Xupeng Geng. 2020. An Effective Maximum Entropy Exploration Approach for Deceptive Game in Reinforcement Learning. *Neurocomputing* (2020).
[34] Jonathan Feng-Shun Lin, Pamela Carreno-Medrano, Mahsa Parsapour, Maram Sakr, and Dana Kulić. 2021. Objective learning from human demonstrations. *Annual Reviews in Control* 51 (2021), 111–129.
[35] Ajay Mandlekar, Danfei Xu, Josiah Wong, Soroush Nasiriany, Chen Wang, Rohun Kulkarni, Li Fei-Fei, Silvio Savarese, Yuke Zhu, and Roberto Martín-Martín. 2021. What matters in learning from offline human demonstrations for robot manipulation. *arXiv preprint arXiv:2108.03298* (2021).
[36] Shaunak A Mehta, Forrest Meng, Andrea Bajcsy, and Dylan P Losey. 2024. StROL: Stabilized and Robust Online Learning from Humans. *IEEE Robotics and Automation Letters* (2024).
[37] Anahita Mohseni-Kabir, Charles Rich, Sonia Chernova, Candace L Sidner, and Daniel Miller. 2015. Interactive hierarchical task learning from a single demonstration. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. 205–212.

[38] Ashvin Nair, Abhishek Gupta, Murtaza Dalal, and Sergey Levine. 2020. Awac: Accelerating online reinforcement learning with offline datasets. *arXiv preprint arXiv:2006.09359* (2020).
[39] Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J Andrew Bagnell, Pieter Abbeel, and Jan Peters. 2018. An algorithmic perspective on imitation learning. *arXiv preprint arXiv:1811.06711* (2018).
[40] Affan Pervez and Dongheui Lee. 2018. Learning task-parameterized dynamic movement primitives using mixture of GMMs. *Intelligent Service Robotics* 11, 1 (2018), 61–78.
[41] Marc Rigter, Bruno Lacerda, and Nick Hawes. 2020. A framework for learning from demonstration with minimal human effort. *IEEE Robotics and Automation Letters* 5, 2 (2020), 2023–2030.
[42] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. 2011. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 627–635.
[43] Maram Sakr, Martin Freeman, H F Machiel Van der Loos, and Elizabeth Croft. 2020. Training Human Teacher to Improve Robot Learning from Demonstration: A Pilot Study on Kinesthetic Teaching. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 800–806.
[44] Maram Sakr, Zexi Jesse Li, H. F. Machiel Van der Loos, Dana Kulix0107;, and Elizabeth A. Croft. 2022. Quantifying Demonstration Quality for Robot Learning and Generalization. *IEEE Robotics and Automation Letters* 7, 4 (2022), 9659–9666. https://doi.org/10.1109/LRA.2022.3191950
[45] Jeff Sauro and James R. Lewis. 2012. *Quantifying the User Experience*. Morgan Kaufmann. https://doi.org/10.1016/C2010-0-65192-3
[46] Aran Sena and Matthew Howard. 2020. Quantifying teaching behavior in robot learning from demonstration. *The International Journal of Robotics Research* 39, 1 (2020), 54–72.
[47] Claude Elwood Shannon. 2001. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review* 5, 1 (2001), 3–55.
[48] Aaron P Shon, Deepak Verma, and Rajesh PN Rao. 2007. Active imitation learning. In *AAAI*. 756–762.
[49] David Silver, J Andrew Bagnell, and Anthony Stentz. 2012. Active learning from demonstration for robust autonomous navigation. In *2012 IEEE International Conference on Robotics and Automation*. IEEE, 200–207.
[50] Jonathan Spencer, Sanjiban Choudhury, Matthew Barnes, Matthew Schmittle, Mung Chiang, Peter Ramadge, and Siddhartha Srinivasa. 2020. Learning from interventions: Human-robot interaction as both explicit and implicit feedback. In *16th Robotics: Science and Systems, RSS 2020*. MIT Press Journals.
[51] Halit Bener Suay, Russell Toris, and Sonia Chernova. 2012. A practical comparison of three robot learning from demonstration algorithm. *International Journal of Social Robotics* 4, 4 (2012), 319–330.
[52] Stone Tao, Arth Shukla, Tse-kai Chan, and Hao Su. 2024. Reverse Forward Curriculum Learning for Extreme Sample and Demo Efficiency. In *The Twelfth International Conference on Learning Representations*.
[53] Stone Tao, Arth Shukla, Tse-kai Chan, and Hao Su. 2024. Reverse forward curriculum learning for extreme sample and demonstration efficiency in reinforcement learning. *arXiv preprint arXiv:2405.03379* (2024).
[54] Agnes Tegen, Paul Davidsson, and Jan A Persson. 2021. Active learning and machine teaching for online learning: a study of attention and labelling cost. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 1215–1220.
[55] Tu Trinh, Haoyu Chen, and Daniel S Brown. 2024. Autonomous assessment of demonstration sufficiency via bayesian inverse reinforcement learning. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 725–733.
[56] Jiakai Zhang and Kyunghyun Cho. 2016. Query-efficient imitation learning for end-to-end autonomous driving. *arXiv preprint arXiv:1605.06450* (2016).
[57] Ruohan Zhang, Dhruva Bansal, Yilun Hao, Ayano Hiranaka, Jialu Gao, Chen Wang, Roberto Martín-Martín, Li Fei-Fei, and Jiajun Wu. 2023. A dual representation framework for robot learning with human guidance. In *Conference on Robot Learning*. PMLR, 738–750.
[58] Xiaojin Zhu. 2015. Machine teaching: An inverse problem to machine learning and an approach toward optimal education. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29.
[59] Xiaojin Zhu, Adish Singla, Sandra Zilles, and Anna N Rafferty. 2018. An overview of machine teaching. *arXiv preprint arXiv:1801.05927* (2018).