

Training Robots with Natural and Lightweight Human Feedback

Erdem Biyik

Thomas Lord Department of Computer Science
University of Southern California

biyik@usc.edu

Abstract—Generalist robot models promise broad applicability across domains but currently require extensive expert demonstrations for task specialization, which is a costly and impractical barrier for real-world deployment. In this article, which summarizes the author’s presentation in the New Faculty Highlights Track of the 39th annual AAAI Conference on Artificial Intelligence,¹ we present algorithms that enable non-expert users to adapt and continually improve robot policies through natural and lightweight feedback modalities, such as preference comparisons, rankings, ratings, natural language, and users’ own demonstrations, combining them with active learning strategies to maximize data-efficiency. We further introduce methods for leveraging real-time human interventions as rich training signals, modeling both their timing and absence to refine policies continually. Our approaches achieve substantial gains in sample-efficiency, adaptability, and user-friendliness, demonstrated across simulated and real-world robotic tasks. By aligning robot learning with how humans naturally teach, we hope to move toward autonomous systems that are more personalized, capable, and deployable in everyday environments.

I. INTRODUCTION

Robots are poised to transform many aspects of everyday life, from helping with home chores and supporting patient care in hospitals to improving agricultural productivity and improving manufacturing. In home environments, for example, robots have the potential to help people with routine tasks like cleaning, laundry, and even self-care, which can especially benefit older adults and individuals with disabilities [2]. In healthcare, robots now perform roles ranging from surgery and rehabilitation to socially assistive care and telepresence [3]. Agricultural robots can handle labor-intensive jobs such as planting, weeding, and harvesting, addressing labor shortages while increasing yields [4]. In industrial settings, collaborative robots (cobots) are introducing new flexibility by working safely alongside humans on assembly lines and in warehouses [5].

Across these domains and many more, the ability to adapt and learn is crucial. Robots must handle diverse tasks and changing conditions, which motivates research in robot learning. One promising avenue is the recent development of generalist robotic models, often in the form of vision-language-action (VLA) policies. These generalist models are trained on extremely large datasets and can often be specialized to be instructed with multimodal inputs (e.g., images and language) to perform specific tasks. For instance, Octo [6] is a transformer-based policy pretrained on hundreds

of thousands of demonstrations, which is then fine-tuned to different robots and tasks. Similarly, Google’s RT-2 VLA model [7] leverages web-scale vision-language pretraining to transfer knowledge into robotic manipulation.

The standard practice for specializing generalist models is to fine-tune them with a large number of expert demonstrations for each new task or environment [6]–[13]. In practice, this means collecting hundreds of high-quality demonstrations (usually teleoperated) per task to achieve reliable performance. This reliance on large expert data poses a major bottleneck outside of laboratory conditions: expert demonstrations are slow and costly to collect [14], especially in unstructured real-world environments with non-technical users. Thus, there is a need for more data-efficient ways to adapt robot policies to specific user needs and contexts.

Many researchers, including our research group, have explored alternative feedback modalities that are easier to obtain from everyday users, such as preference comparisons [15]–[17] and rating-based feedback (e.g., bad vs. good vs. very good) [18]–[21]. Rather than providing a full demonstration of a task, a human can simply choose which of two robot behaviors is better, or rate the behavior. These forms of feedback lower the bar for user participation. For example, it is often easier for a person to select a preferred trajectory than to kinesthetically demonstrate the task from scratch [22]. Such feedback modalities have been quite useful in other domains that adopted large pretrained models, and we expect increasingly more interest and applications in robotics, too. We present and discuss our contributions in this direction in Section II.

The challenge, however, is that these lightweight feedback modalities tend to provide information in small increments, meaning the robot may require an even larger number of interactions than demonstrations to learn an effective policy [22], [23]. Therefore, preference- and rating-based feedback can reduce the upfront burden of demonstration collection, but often at the cost of needing many more rounds of interaction, which is itself impractical for non-expert end-users.

Moreover, none of these feedback types is natural: people typically teach each other by using their own demonstrations (without “teleoperating” their partner) or by talking in natural language [24]–[29]. In Section III, we are motivated by this mismatch between human-human and human-robot interactions. We ask: *can we develop algorithms that are accessible to non-expert end-users by enabling robot learning from easy*

¹An abstract of the talk was published in [1].

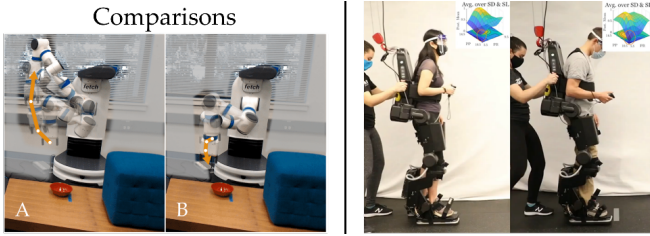


Fig. 1: (left) We developed algorithms to learn from humans’ preference comparisons [17], (right) which enabled learning the gait preferences of the lower-body exoskeleton users [21].

and natural types of human feedback, such as natural language and humans’ own demonstrations? These algorithms will enable everyday people to fine-tune generalist models with ease to achieve specific tasks.

Finally, once robots are deployed in real-world environments, they will inevitably make mistakes. But each mistake, if handled well, is an opportunity for the robot to learn. Consider a robot that is helping in a kitchen: if it reaches for the wrong object or is about to spill something, the human user will instinctively intervene or even correct the robot. Traditionally, such interventions are viewed as failures to be minimized, but an emerging perspective treats them as valuable feedback signals [30]–[35]. When a user presses an emergency stop or physically guides the robot away from an error, the event carries information about undesirable actions or states [36]. Prior works have incorporated these signals in relatively heuristic ways, e.g., by assigning a negative reward to the robot’s policy whenever the human intervenes [34]. Another approach is to impose safety constraints on the robot’s model upon intervention: for instance, an algorithm might treat the trajectory segment that led to an intervention as violating a constraint and adjust the policy to downweight that behavior [35]. These intervention-based techniques have shown improvements in sample-efficiency and safety. However, they often treat the human input in a simplistic way (e.g., every intervention is a uniform penalty), losing nuance about how the human corrected the robot or what the human’s goal was.

Moving forward, *there is a need for a more principled modeling of human interventions and corrections*. Using such computational models, the robot could extract far more information per feedback than a binary penalty. In essence, rather than hard-coding interventions as negative rewards or constraints, the robot should learn from interventions in a rich way, inferring the underlying intended lesson and generalizing it to future situations, which promises dramatically higher data-efficiency. We present some preliminary results in this direction in Section IV.

In summary, generalist robotics models stand to impact a broad range of fields, but we must enable them to learn effectively from everyday human feedback. Our research thus points toward unifying various feedback types in a way that leverages the strengths of each. By modeling the human as a fallible but informative teacher and the robot as an active learner, we can begin to bridge the gap between *human teaching* and *robot learning*. Ultimately, our research aims to allow robots to continually learn from the natural

interactions that occur as people use them, making robots more personalized and adaptable [37].

II. LEARNING FROM PREFERENCE COMPARISONS

A core research direction we have worked on is preference-based learning. More recently known as reinforcement learning from human feedback (RLHF) [16], preference-based learning algorithms have been the standard technique for aligning foundation models [38] with human preferences and societal values [39]. In this framework, an AI system presents (at least) two options to a human user and asks “which one do you prefer?” This is visualized in Figure 1(left). The human’s response to this question reveals information about the objective function the AI system should optimize. For example, we denote a pairwise comparison question between two options (e.g., robot trajectories) τ_A and τ_B as $S = \{\tau_A, \tau_B\}$. Without loss of generality, when the user responds with $q = \tau_A$, a Bayesian update is performed for the parameters ω of the reward function the AI agent is trying to learn:

$$P(\omega \mid \mathcal{D}, S, q) \propto P(\omega) \prod_{(S, q) \in \mathcal{D}} P(q \mid \omega, S)$$

where \mathcal{D} is the data collected from the human up to the current iteration. On the right-hand side, the first term is simply the prior belief about the parameters ω . The second term is the likelihood, and different choice models are adopted in the literature, such as Bradley-Terry model [40] or Thurstonian model [41]:

$$\text{Bradley-Terry: } P(q \mid \omega, S) \propto \exp R_\omega(q)$$

$$\text{Thurstonian: } P(q \mid \omega, S) \propto R_\omega(q) + \epsilon - \min_{\bar{q} \in S} R_\omega(\bar{q})$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise added independently to different options $q' \in S$. In addition to these *noisily optimal* models, we also studied models from behavioral economics that aim to better capture human decision-making by incorporating humans’ biases, such as cumulative prospect theory [42], [43].

Regardless of the likelihood model used, this preference-based learning framework allows the human to give feedback to the system without controlling it themselves, i.e., without giving a demonstration. It enables humans to train or adapt AI systems simply by stating a choice between the presented options. Therefore, preference feedback is considered easy-to-collect.

In our research, we developed a family of algorithms that convert preference feedback into reward functions and control policies for robots. We have formalized pairwise preferences, rankings, scalar ratings, and integrated them with other feedback modalities, like demonstrations, in a unified Bayesian learning framework [44], [45].

A. Pairwise preferences for reward learning

For preference-based learning to be useful in robotics, it needs to be both data-efficient, so that users will not have to interact with the robot for thousands of preference data samples to teach it a single task, and user-friendly, so that users will be able to train robots by reliably indicating their

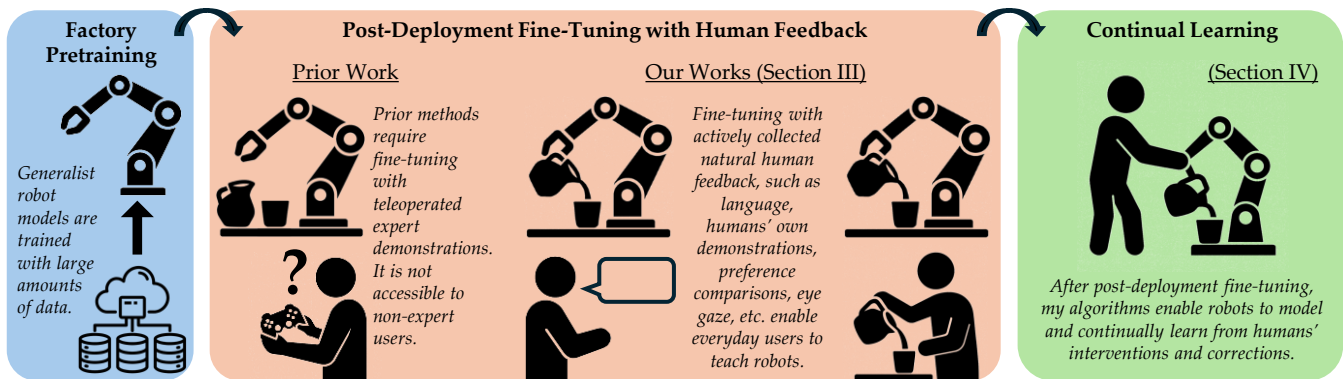


Fig. 2: The works we present in Section III and Section IV discuss different ways non-expert humans can train robots post-deployment. As opposed to the common practice, we avoid fine-tuning with expert demonstrations, which are not accessible to everyday people.

preferences.

To increase both data-efficiency and user-friendliness, we developed active learning methods that enabled the robot to intelligently choose which preference comparison question to ask to the user. Specifically, our information gain based method [23], where the robot asks questions that maximize the mutual information between the user’s response and the parameters to be learned (e.g., the weights of the reward function), uses the following acquisition function:

$$S^* = \underset{S}{\operatorname{argmax}} I(q; \omega \mid \mathcal{D}, S).$$

By maximizing the mutual information in this way, our method has led to $10\times$ improvement in data-efficiency over the baselines at that time, e.g., volume removal [15]. This formulation also enabled us to derive an optimal stopping rule and to incorporate users’ “about the same” responses [46]. Our information gain based active querying method for preferences is still widely used as the state-of-the-art, and our open-source library APReL [47] makes it convenient for newcomers to start using different active learning algorithms with different forms of feedback and likelihood functions. We later extended our preference-based learning and active querying algorithms to arbitrary objective functions (rather than the specific case of minimizing uncertainty about the parameters ω) [48], to the cases where prior can be constructed with expert demonstrations [17], and to nonlinear reward functions, like neural networks [49] or Gaussian processes [50], [51].

These improvements over data-efficiency and functional forms of reward functions enabled us to deploy our preference-based algorithms in many different application domains, including minimizing congestion in traffic networks by setting prices in ride-hailing services or taxi fares after learning users’ price-latency tradeoff [52]–[54], recommendation systems [55], and a lower-body exoskeleton platform where we leveraged users’ preference data to optimize their comfort and safety [21] (see Figure 1(right)).

While the use of active learning techniques like information gain maximization improves data-efficiency, they require solving an optimization problem for each and every query, incurring a large computation overhead. Thus, we developed batch-mode active preference learning to generate

diverse, informative sets of comparison queries at once, enabling faster wall-clock progress without redundant questions [56]. We used determinantal point processes (DPP) [57] to pick batches that balance information gain with diversity, and showed on multiple robotic simulations that these batches converge quickly while dramatically reducing query-generation time [58], [59].

B. Beyond pairwise comparisons

Our research broadened feedback beyond pairwise comparisons in multiple major directions.

1) Multimodal rewards from rankings

When data come from multiple people (or a single user with multiple goals) preferences are inherently multimodal. We formulated reward learning as a mixture of Plackett–Luce ranking models [60] and extended our active information-gain objective for ranking queries [61].

2) Hierarchical queries for dynamic reward functions

People’s preferences change over time depending on their interactions with the world and the AI system. To capture this, we developed hierarchical preference queries where scenarios follow each other based on the user’s choices [62]. This enabled us to learn dynamically changing reward functions from preferences.

3) Preferences with magnitude

After observing that the presence of “about equal” option improved both data-efficiency and user-friendliness, we developed a user interface that allows users to not only indicate their preference but also signal the magnitude of their preference, i.e., how much more they prefer one option over the other [63]. We also developed a likelihood model, i.e., human choice model, for such feedback. In this way, we significantly improved the information bandwidth of preference feedback.

4) Attribute preferences

It is often useful for the system to know *why* a user preferred one option over the other. To this end, we developed models of learning from *attribute queries* where the system asks if the user prefers more or less of an attribute (e.g., higher or lower speed) [55], decreasing the amount of required interactions with the user by further improving the information bandwidth.

Another limitation of standard preference-based feedback for robotics is its requirement for the user to watch multiple trajectories of the system before deciding which one they prefer. However, a more natural way to convey preferences is by using language. As humans, we simply give instructions or corrections in natural language, e.g., “cut the künefe in larger slices.” Enabled by the advances in language models, we formulated reward learning problem as a machine translation problem where translation happens between trajectory and language spaces. This led to preference learning algorithms that are almost $2\times$ faster in wall-clock time than even learning from pairwise comparisons as the users now need to observe only a single trajectory [64]. Besides, the fact that the users themselves are choosing which aspect of the robot’s behavior they give feedback about (e.g., slice size but not the amount of syrup) further improved data-efficiency as it also carries information about their preferences.

5) Learning from ratings

Finally, we developed algorithms that enable preference data to be used with ratings, numerical scores humans assign to options, which are abundant in some applications like recommendation systems. Extending our information gain based active querying method to this mixed ratings-and-preferences setup allowed us to learn gait preferences of lower-body exoskeleton users where data-efficiency is extremely important [21] (see Figure 1(right)).

III. LEARNING FROM NATURAL TYPES OF FEEDBACK

While different forms of preference data paired with active querying algorithms present a convenient way for users to train/adapt robots, their information bandwidth is extremely limited compared to expert demonstrations collected via teleoperation. Expert demonstrations, on the other hand, are not accessible to everyday users who are not technical experts (see Figure 2).

To remedy these, we investigate how robots can learn from the two most common and natural types of human feedback [24]–[29]: natural language, and users performing the task on their own without the robot. We develop techniques to generate training signals for the robots from language by tapping into pretrained models, or synthesize robot behaviors from humans’ own demonstrations by retrieving relevant robot skills. Ultimately, we aim to develop robots that learn from multimodal human feedback, which include traditional feedback types like teleoperated demonstrations and preferences, as well as natural types such as language, humans’ own demonstrations, gestures, eye gaze, interventions, etc.

To achieve robots that are useful in real-life environments, such as homes, schools, small businesses, they need to be easily adaptable by everyday people. Put another way, people who are not experts in robotics must be able to teach generalist robots new skills in a convenient way. The most important component that will make this possible but the current literature is missing is robot learning algorithms that accept natural human feedback, i.e., feedback humans naturally adopt while interacting with and teaching each other, as input.

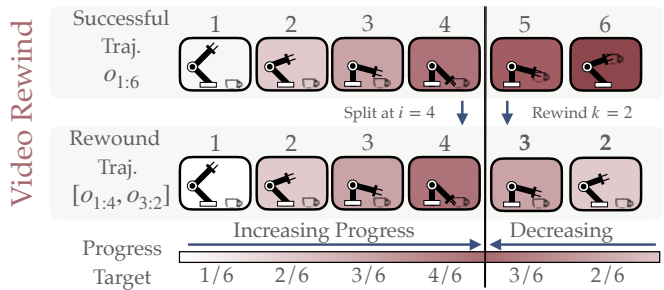


Fig. 3: ReWiND [68] splits a demo at intermediate timestep i into forward/reverse sections. Here, the forward section shows the robot approaching the cup; the reverse section (o_{i-1}, o_{i-2}, \dots) resembles dropping it.

First, we ask: how do we have robots learn from open-ended natural language instructions?

A. Learning from Natural Language

Our prior work and other state-of-the-art methods use large pretrained models, e.g., video-language models in RoboCLIP [65] or vision-language models in RL-VLM-F [66] and IMPACT [67], to generate reward signals or cost functions for robots. Since these models were trained mainly on human data, i.e., videos of humans and not robots, these approaches require these pretrained models to generalize from human data to robot data.

Here, we discuss our recent method, ReWiND, which attempts to solve this problem by fine-tuning a reward model only by using existing demonstration videos, without requiring new demonstrations of the target task [68]. The key idea is to build a language-conditioned reward function from a small, fixed set of demonstrations and then use it to guide both offline pretraining and online adaptation. This begins with a dataset in the target environment containing only a handful of demonstrations per task. Importantly, these tasks can be different from the target task. From them, we train a reward model that predicts fine-grained task progress for a video trajectory given a natural language instruction. Unlike prior methods that rely on sparse success signals (e.g., RoboCLIP [65]) or relative comparisons (e.g., RL-VLM-F [66]), ReWiND directly regresses to normalized per-frame progress, which naturally yields dense reward signals. To help the reward model generalize and remain robust, we incorporate additional diverse data from the Open X-Embodiment dataset [69], augment the demonstrations with synthetic language paraphrases, and create artificial “failure” examples via a video rewinding procedure that simulates the robot making and recovering from mistakes as shown in Figure 3.

Architecturally, the reward model uses frozen pretrained vision and language encoders: DINOv2 for images [70] and a compact MiniLM variant for text [71], feeding into a lightweight cross-modal sequential transformer. By freezing the encoders, we avoid overfitting to the small in-domain dataset while leveraging the generality of large-scale pretraining. A positional embedding is applied only to the first frame to capture necessary temporal cues without allowing the model to “cheat” by relying solely on frame indices. During training, we mix real demonstration tra-

jectories with mismatched instruction–video pairs (assigned zero progress) and rewind sequences (assigned decreasing progress), thereby teaching the model to assign appropriate low rewards to off-task or failing behaviors it may encounter during online learning.

Once trained, the reward function is used to label the original demonstration data with dense progress-based rewards, and we pretrain a language-conditioned policy using offline reinforcement learning. This step gives the policy a reasonable initialization for interacting in the environment, even on tasks it has not yet seen. For new tasks, the robot needs only a language description; it executes the pretrained policy, receives reward labels from the frozen reward model, and fine-tunes online. Because the reward model has learned to handle diverse instructions, unseen objects, and failure modes, it can guide the policy toward successful behavior in a sample-efficient manner.

Empirically, ReWiND’s design yields reward functions that align more consistently with task progress, better distinguish between levels of success in policy rollouts, and remain stable across varied language inputs compared to existing baselines. In simulated Meta-World experiments [72], policies fine-tuned with ReWiND achieve much higher success rates on unseen tasks in low data regime, and in real-world bimanual manipulation, an hour of online training with ReWiND improved pretrained policies by a factor of five. These results suggest that carefully modeling progress, incorporating synthetic diversity, and grounding learning in both limited in-domain and broad out-of-domain data can make reward functions an effective bridge between natural language task descriptions and practical robot learning without the constant need for new demonstrations.

B. Learning from Humans’ Own Demonstrations

While fine-tuning online via natural language instructions creates an easy interface for non-expert users, it requires either real-world reinforcement learning training or closing the sim-to-real gap, both of which are difficult problems that have been around for many years. Imitation learning, on the other hand, is significantly more data-efficient compared to RL [73]–[80]. Indeed, humans often teach one another not through verbal instruction alone, but by physically performing the task themselves. In human-robot interaction, this form of feedback has been underutilized due to the embodiment gap between human and robot morphologies, which makes direct imitation difficult. To overcome this, we developed an approach that translates human demonstrations into robot behaviors by leveraging retrieval-based imitation from existing pre-deployment robot data.

As an example, suppose that a user wants to teach their robot how to put a tablecloth on a table. The user may perform this task themselves while the robot is simply observing with its cameras. While it is difficult for the robot to directly imitate the human due to the embodiment gap, it may retrieve relevant trajectories from its memory to learn the task efficiently via imitating that specific task from its memory. For example, if its pre-deployment data include

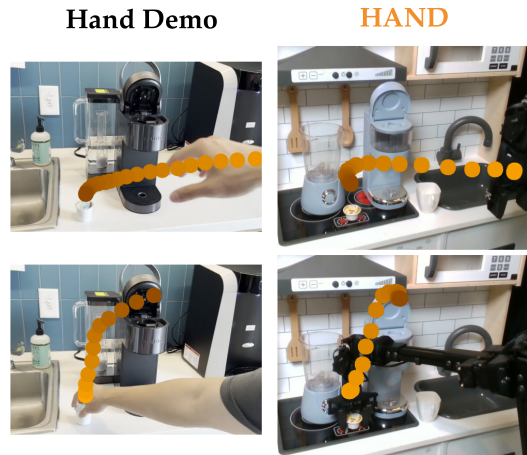


Fig. 4: Visualizations from [81] that shows HAND’s top sub-trajectory match on two out-of-domain demonstrations recorded from an iPhone camera, showing approaching a K-Cup and putting it into the machine.

some (sub-)trajectories of making bed, these demonstrations carry important information for the target task of putting tablecloth due to the similarities between the tasks. In our recent work HAND [81], we investigate filtering and retrieval methods to identify tasks similar to the target task.

Specifically, HAND is a method we developed for rapidly adapting robots to new manipulation tasks using only a single human hand demonstration, without requiring teleoperated robot demonstrations. The core idea is to leverage large, task-agnostic play datasets and retrieve a subset of trajectories that resemble the human-provided demonstration. To do this, HAND first tracks the motion of the human hand in the demonstration video using a simple 2D point-tracking pipeline, producing a relative hand path that abstracts away starting position and background details. The same process is applied to the robot’s gripper in the play dataset. To avoid retrieving similar motions in completely irrelevant settings, a visual filtering step uses features from a pretrained DINOv2 model [70] to retain only robot trajectories involving similar environments and objects. Finally, HAND compares the filtered candidates to the demonstration path using subsequence dynamic time warping [82], retrieving those with the most similar motion patterns as shown in Figure 4.

Once relevant sub-trajectories are retrieved, HAND fine-tunes a policy that was pretrained on the entire play dataset. This fine-tuning is done with parameter-efficient LoRA adapters [83] inserted into the transformer-based policy, allowing the model to adapt to the demonstrated task with only a small fraction of its weights updated. To emphasize the most relevant retrievals, HAND weights each imitation loss term according to the similarity score between the retrieved trajectory and the hand demonstration. This ensures that the fine-tuning process focuses on the closest motion matches while still benefiting from some diversity in the training examples. By using retrieval instead of collecting new robot data, HAND reduces both data collection time and the technical skill needed from users.

In simulation via CALVIN benchmark [84] and on a real WidowX-250 arm, HAND achieves significantly higher

success rates than retrieval baselines that rely solely on visual similarity or optical flow, and it remains robust even when hand demonstrations are recorded in different environments. Real-world trials show that HAND can learn complex, long-horizon tasks in under four minutes from a single hand demonstration, with demonstrations being about five times faster to collect than robot teleoperation data. This approach makes it possible for non-expert users to adapt a generalist robot to new, specific tasks in real time, using the kind of natural, example-based teaching that people already employ with one another.

IV. CONTINUAL LEARNING FROM INTERVENTIONS

Even after a robot is deployed and adapted to its specific tasks via preference data (Section II) or natural human feedback (Section III), it will inevitably encounter new situations, make occasional mistakes, or face uncertainties the user wishes to preempt. Rather than viewing these instances as failures, we should treat them as key learning opportunities for the robot. In real-world settings, non-expert users often correct the robot in intuitive ways: stopping it before a mistake occurs or kinesthetically intervening after an error. These interventions and corrections are rich, underutilized signals that can drive continual learning.

This section focuses on converting such post-deployment interactions into a framework for continual robot learning. We develop computational models and algorithms that go beyond treating corrections as additional demonstration data. Our goal is to ensure that every corrective interaction a user makes (or even does not make, as long as the robot knows the human has the chance to intervene) helps the robot become more aligned with the user’s needs.

Specifically, we developed both robot-gated and human-gated algorithms: in robot-gated algorithms, robots proactively pause and query the user with a question during task execution. For example, we extended our preference-based learning works to this setting where the robot asked the user between alternative actions [85]. By employing expected value of information (EVOI) [86], [87], we optimized both *when* the robot should pause to ask a question and *what* action alternatives it should present.

Unfortunately, this method is mostly limited to the settings where the robot’s action alternatives can be visualized in a way that is easily understandable by the user. And akin to preference comparisons, this online feedback also has very low information bandwidth. Similar to how demonstrations are more informative than preferences, human interventions where the user takes over the control are more informative than robot asking preferences between action alternatives.

To this end, we developed human-gated methods to learn from preemptive interventions. In this framework, the robot operates with its policy and the human can intervene any time at will to take over the control, which can be used as a training signal. Our method named MILE [88] has shown that these feedback signals encode nuanced information, particularly in their *timing* and *context*. For example, the moment at which a human intervenes reveals implicit thresholds for

failure and may signal critical decision boundaries in the policy space.

Existing methods for learning from human corrections typically treat interventions as simple labels, either negative rewards [34] or surrogate demonstrations [32], without modeling the nuanced reasoning behind *when* and *how* users intervene or correct the robot. This approach overlooks a critical insight: *interventions and corrections are deeply contextual and shaped by the human’s understanding of the robot’s behavior and the perceived risk of failure*. For example, a user may preemptively stop the robot not because an error has occurred, but because they anticipate one. Prior work lacks a principled model of these dynamics, resulting in inefficient learning and limited generalization. Our key insight is to *treat interventions not as isolated data points, but as probabilistic, strategically chosen signals that reflect users’ mental models of the robot’s capabilities and intent*.

To model how robots can learn from preemptive human interventions, we developed a probabilistic framework that captures the decision process behind such interventions. Let $Q_\theta(s, a)$ be the true but unknown state-action value function of the task and π_θ the corresponding policy of the robot. We model the human’s belief about the robot’s objective and the policy using a pair $(\hat{Q}_\xi(s, a), \hat{\pi}_\xi)$. This model allows us to express preemptive interventions as a discrete choice problem: given a state s , will the human prefer to let the robot continue acting, or take over control, e.g., by physical interaction?

We quantified the probability of a human intervention as follows. The human has an action $a \sim \pi_\theta(\cdot | s)$ in their mind about what the robot should do, and compares it to what they believe the robot will do $a' \sim \hat{\pi}_\xi(\cdot | s)$. Then, the intervention probability is:

$$p(\nu = 1 | s) = \mathbb{E}_{a \sim \pi_\theta(\cdot | s)} \left[\Phi \left(\mathbb{E}_{a' \sim \hat{\pi}_\xi(\cdot | s)} [Q_\theta(s, a) - Q_\theta(s, a') - c] \right) \right], \quad (1)$$

where Φ is the cdf of the standard normal distribution, and c represents the cost of intervening (e.g., physical or cognitive effort). This model encodes the idea that the human will only intervene when the expected value of their own action exceeds that of the robot by more than the intervention cost. Because this model is fully differentiable with respect to θ , which represents optimal parameters for the task, we conveniently used it to optimize the robot’s policy π_θ . Under Boltzmann rationality assumption, we further wrote Equation (1) in terms of the policies instead of Q -functions, thereby eliminating the need for policy optimization, which increased the time-efficiency and feasibility of training (see Figure 5).

This computational intervention model represents a significant departure from traditional approaches that treat interventions merely as surrogate demonstrations. Rather than using human interventions only to overwrite the robot’s actions with direct labels, this model enables robots to learn not only from when the human intervenes, but also from when they choose not to. In situations where the robot acts autonomously and the human, who is present and able to

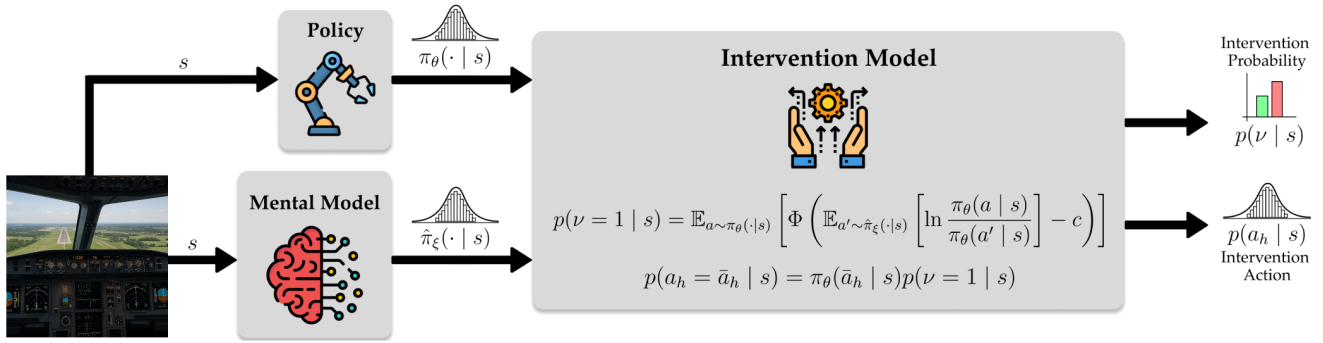


Fig. 5: Starting from an initial policy π_θ , MILE [88] jointly trains the mental model $\hat{\pi}_\xi$, and the policy using our computational model of preemptive interventions. \bar{a}_h denotes what the human would do if they were controlling the system, and a_h is their realized action, which may be a non-intervention.

intervene, opts not to intervene, the robot can infer that its behavior was within the acceptable range of performance (Q -value) according to the mental model of the human. Thus, the absence of intervention becomes implicit positive feedback, helping to refine the robot’s policy and expand its confidence in similar situations. This richer interpretation of human interventions enabled more data-efficient, context-aware, and continuous adaptation over time.

V. CONCLUSION

In conclusion, advancing robot learning hinges on making adaptation accessible, efficient, and natural for everyday users. By unifying diverse feedback modalities, from preference comparisons and natural language to human demonstrations and real-time interventions, our work attempts to bridge the gap between how humans naturally teach and how robots currently learn. Through active learning strategies, multimodal feedback integration, and principled modeling of human input, our approaches dramatically improve data-efficiency and user-friendliness. Ultimately, these methods pave the way for robots that not only adapt to novel tasks in real-world environments but also evolve continuously through intuitive human interaction, fostering more capable, personalized, and trustworthy autonomous systems.

Moving forward, we envision that future robots and autonomous systems will need to tap into a much more diverse set of human feedback to address data limitations [37]. To this end, our research group now works on incorporating *natural* and *almost-no-effort* human feedback into robot learning pipelines. For example, we develop algorithms that track humans’ eye gaze to understand the causal and salient parts of the robot’s environment [89], [90], improving the data-efficiency of imitation learning for more than 25% with no extra human effort. We believe, in addition to gaze, hand gestures [91], facial expressions [92], tone of voice in speech, and many more implicit cues are underutilized feedback signals in robot learning which may be useful as we deploy robots in real-world.

ACKNOWLEDGMENTS

The author would like to thank Dorsa Sadigh and the members of the ILIAD group at Stanford University, who he collaborated on works presented in Section II. The author

also thanks the members of Lira Lab at the University of Southern California, who led the projects presented in Sections III and IV. This work was partially funded by Airbus Institute for Engineering Research (AIER). The author has no conflict of interest to declare.

REFERENCES

- [1] E. Bıyık, “Efficient robot learning via interaction with humans,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 2025, pp. 28 700–28 700.
- [2] C.-A. Smarr, A. Prakash, J. M. Beer, T. L. Mitzner, C. C. Kemp, and W. A. Rogers, “Older adults’ preferences for and acceptance of robot assistance for everyday living tasks,” in *Proceedings of the human factors and ergonomics society annual meeting*, Sage Publications Sage CA: Los Angeles, CA, vol. 56, 2012, pp. 153–157.
- [3] A. A. Morgan, J. Abdi, M. A. Syed, G. E. Kohen, P. Barlow, and M. P. Vizcaychipsi, “Robots in healthcare: A scoping review,” *Current robotics reports*, vol. 3, no. 4, pp. 271–280, 2022.
- [4] M. Spagnuolo, G. Todde, M. Caria, N. Furnitto, G. Schillaci, and S. Failla, “Agricultural robotics: A technical review addressing challenges in sustainable crop production,” *Robotics*, vol. 14, no. 2, p. 9, 2025.
- [5] E. Montini, F. Daniele, L. Agbomemewa, M. Confalonieri, V. Cutrona, A. Bettoni, P. Rocco, and A. Ferrario, “Collaborative robotics: A survey from literature and practitioners perspectives,” *Journal of Intelligent & Robotic Systems*, vol. 110, no. 3, p. 117, 2024.
- [6] Octo Model Team, D. Ghosh, H. Walke, K. Pertsch, K. Black, O. Mees, S. Dasari, J. Hejna, C. Xu, J. Luo, T. Kreiman, Y. Tan, L. Y. Chen, P. Sanketi, Q. Vuong, T. Xiao, D. Sadigh, C. Finn, and S. Levine, “Octo: An open-source generalist robot policy,” in *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, 2024.
- [7] B. Zitkovich, T. Yu, S. Xu, P. Xu, T. Xiao, F. Xia, J. Wu, P. Wohlhart, S. Welker, A. Wahid, *et al.*, “Rt-2: Vision-language-action models transfer web knowledge to robotic control,” in *Conference on Robot Learning*, PMLR, 2023, pp. 2165–2183.
- [8] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. P. Foster, P. R. Sanketi, Q. Vuong, *et al.*, “Openvla: An open-source vision-language-action model,” in *8th Annual Conference on Robot Learning*, Nov. 2024.
- [9] A.-C. Cheng, Y. Ji, Z. Yang, X. Zou, J. Kautz, E. Bıyık, H. Yin, S. Liu, and X. Wang, “Navila: Legged robot vision-language-action model for navigation,” in *Proceedings of Robotics: Science and Systems (RSS)*, Jun. 2025.
- [10] Y. Ma, Z. Song, Y. Zhuang, J. Hao, and I. King, “A survey on vision-language-action models for embodied ai,” *arXiv preprint arXiv:2405.14093*, 2024.
- [11] Y. Li, Y. Deng, J. Zhang, J. Jang, M. Memmel, R. Yu, C. R. Garrett, F. Ramos, D. Fox, A. Li, *et al.*, “Hamster: Hierarchical action models for open-world robot manipulation,” in *The Thirteenth International Conference on Learning Representations*, 2025.
- [12] T. Van Vo, T. Q. Nguyen, K. M. Nguyen, D. H. M. Nguyen, and M. N. Vu, “Refinevla: Reasoning-aware teacher-guided transfer fine-tuning,” *arXiv preprint arXiv:2505.19080*, 2025.

- [13] C. Xu, Q. Li, J. Luo, and S. Levine, "Rldg: Robotic generalist policy distillation via reinforcement learning," in *Proceedings of Robotics: Science and Systems (RSS)*, 2025.
- [14] S. Reed, K. Zolna, E. Parisotto, S. G. Colmenarejo, A. Novikov, G. Barth-maroon, M. Giménez, Y. Sulsky, J. Kay, J. T. Springenberg, et al., "A generalist agent," *Transactions on Machine Learning Research*, 2022.
- [15] D. Sadigh, A. D. Dragan, S. S. Sastry, and S. A. Seshia, "Active preference-based learning of reward functions," in *Proceedings of Robotics: Science and Systems (RSS)*, Jul. 2017.
- [16] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," *Advances in neural information processing systems*, vol. 30, 2017.
- [17] E. Biyik, D. P. Losey, M. Palan, N. C. Landolfi, G. Shevchuk, and D. Sadigh, "Learning reward functions from diverse sources of human feedback: Optimally integrating demonstrations and preferences," *The International Journal of Robotics Research (IJRR)*, 2021.
- [18] W. B. Knox and P. Stone, "Tamer: Training an agent manually via evaluative reinforcement," in *2008 7th IEEE international conference on development and learning*, IEEE, 2008, pp. 292–297.
- [19] J. MacGlashan, M. K. Ho, R. Loftin, B. Peng, G. Wang, D. L. Roberts, M. E. Taylor, and M. L. Littman, "Interactive learning from policy-dependent human feedback," in *International conference on machine learning*, PMLR, 2017, pp. 2285–2294.
- [20] D. Arumugam, J. K. Lee, S. Saskin, and M. L. Littman, "Deep reinforcement learning from policy-dependent human feedback," *arXiv preprint arXiv:1902.04257*, 2019.
- [21] K. Li, M. Tucker, E. Biyik, E. Novoseller, J. W. Burdick, Y. Sui, D. Sadigh, Y. Yue, and A. D. Ames, "Roial: Region of interest active learning for characterizing exoskeleton gait preference landscapes," in *International Conference on Robotics and Automation (ICRA)*, May 2021.
- [22] M. Palan, G. Shevchuk, N. Charles Landolfi, and D. Sadigh, "Learning reward functions by integrating human demonstrations and preferences," in *Robotics: Science and Systems*, 2019.
- [23] E. Biyik, M. Palan, N. C. Landolfi, D. P. Losey, and D. Sadigh, "Asking easy questions: A user-friendly approach to active reward learning," in *Proceedings of the 3rd Conference on Robot Learning (CoRL)*, 2019.
- [24] S. Ellis and B. Rogoff, "The strategies and efficacy of child versus adult teachers," *Child development*, pp. 730–735, 1982.
- [25] T. R. Sumers, M. K. Ho, R. D. Hawkins, and T. L. Griffiths, "Show or tell? exploring when (and why) teaching with language outperforms demonstration," *Cognition*, vol. 232, p. 105 326, 2023.
- [26] D. Yu, N. Goodman, and J. Mu, "Characterizing tradeoffs between teaching via language and demonstrations in multi-agent systems," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 45, 2023.
- [27] R. T. Nabavi, "Bandura's social learning theory & social cognitive learning theory," *Theory of Developmental Psychology*, vol. 1, no. 1, pp. 1–24, 2012.
- [28] A. Bandura and R. H. Walters, *Social learning theory*. Prentice hall Englewood Cliffs, NJ, 1977, vol. 1.
- [29] A. Yu and R. Mooney, "Using both demonstrations and language instructions to efficiently learn robotic tasks," in *International Conference on Learning Representations (ICLR)*, 2023.
- [30] J. Spencer, S. Choudhury, M. Barnes, M. Schmittle, M. Chiang, P. Ramadge, and S. Srinivasa, "Expert intervention learning: An online framework for robot learning from explicit and implicit human feedback," *Autonomous Robots*, pp. 1–15, 2022.
- [31] A. Mandlkar, D. Xu, R. Martín-Martín, Y. Zhu, L. Fei-Fei, and S. Savarese, "Human-in-the-loop imitation learning using remote teleoperation," *arXiv preprint arXiv:2012.06733*, 2020.
- [32] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer, "Hg-dagger: Interactive imitation learning with human experts," in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 8077–8083.
- [33] H. Liu, S. Nasiriany, L. Zhang, Z. Bao, and Y. Zhu, "Robot learning on the job: Human-in-the-loop autonomy and learning during deployment," *The International Journal of Robotics Research*, p. 02 783 649 241 273 901, 2022.
- [34] J. Luo, P. Dong, Y. Zhai, Y. Ma, and S. Levine, "Rlif: Interactive imitation learning as reinforcement learning," in *The Twelfth International Conference on Learning Representations*, 2024.
- [35] J. Spencer, S. Choudhury, M. Barnes, M. Schmittle, M. Chiang, P. Ramadge, and S. Srinivasa, "Learning from interventions: Human-robot interaction as both explicit and implicit feedback," in *Proceedings of Robotics: Science and Systems*, MIT Press Journals, 2020.
- [36] H. J. Jeon, S. Milli, and A. Dragan, "Reward-rational (implicit) choice: A unifying formalism for reward learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 4415–4426, 2020.
- [37] K. Baraka, T. K. Faulkner, E. Biyik, B. Serena, M. Chetouani, D. H. Grollman, A. Saran, E. Senft, S. Tulli, A.-L. Vollmer, A. Andriella, H. Beierling, T. Horter, J. Kober, I. Sheidlower, M. E. Taylor, S. V. Waveren, and X. Xiao, "Human-interactive robot learning: Definition, challenges, and recommendations," 2025.
- [38] R. Bommasani et al., "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.
- [39] S. Casper, X. Davies, C. Shi, T. K. Gilbert, J. Scheurer, J. Rando, R. Freedman, T. Korbak, D. Lindner, P. Freire, T. Wang, S. Marks, C.-R. Segerie, M. Carroll, A. Peng, P. Christoffersen, M. Damani, S. Slocum, U. Anwar, A. Siththaranjan, M. Nadeau, E. J. Michaud, J. Pfau, D. Krasheninnikov, X. Chen, L. Langosco, P. Hase, E. Biyik, A. Dragan, D. Krueger, D. Sadigh, and D. Hadfield-Menell, "Open problems and fundamental limitations of reinforcement learning from human feedback," *Transactions on Machine Learning Research (TMLR)*, 2023.
- [40] R. A. Bradley and M. E. Terry, "Rank analysis of incomplete block designs: I. the method of paired comparisons," *Biometrika*, vol. 39, no. 3/4, pp. 324–345, 1952.
- [41] L. Thurstone, "A law of comparative judgment," *Psychological Review*, vol. 34, no. 4, pp. 273–286, 1927.
- [42] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.
- [43] M. Kwon, E. Biyik, A. Talati, K. Bhasin, D. P. Losey, and D. Sadigh, "When humans aren't optimal: Robots that collaborate with risk-aware humans," in *ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2020.
- [44] E. Biyik, "Learning preferences for interactive autonomy," Ph.D. dissertation, EE Department, Stanford University, 2022.
- [45] E. Biyik, "Learning from humans for adaptive interaction," in *The 17th Annual Human-Robot Interaction Pioneers Workshop (HRI Pioneers)*, Mar. 2022.
- [46] K. Krishnan, "Incorporating thresholds of indifference in probabilistic choice models," *Management science*, vol. 23, no. 11, pp. 1224–1233, 1977.
- [47] E. Biyik, A. Talati, and D. Sadigh, "Aprel: A library for active preference-based reward learning algorithms," in *17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Mar. 2022.
- [48] E. Ellis, G. R. Ghosal, S. J. Russell, A. Dragan, and E. Biyik, "A generalized acquisition function for preference-based reward learning," in *International Conference on Robotics and Automation (ICRA)*, 2024.
- [49] S. M. Katz, A. Maleki, E. Biyik, and M. J. Kochenderfer, "Preference-based learning of reward function features," *arXiv preprint arXiv:2103.02727*, 2021.
- [50] E. Biyik, N. Huynh, M. J. Kochenderfer, and D. Sadigh, "Active preference-based gaussian process regression for reward learning," in *Proceedings of Robotics: Science and Systems (RSS)*, Jul. 2020.
- [51] E. Biyik, N. Huynh, M. J. Kochenderfer, and D. Sadigh, "Active preference-based gaussian process regression for reward learning and optimization," *The International Journal of Robotics Research*, vol. 43, no. 5, pp. 665–684, 2024.
- [52] E. Biyik, D. A. Lazar, D. Sadigh, and R. Pedarsani, "The green choice: Learning and influencing human decisions on shared roads," in *Proceedings of the 58th IEEE Conference on Decision and Control (CDC)*, Dec. 2019.
- [53] M. Beliaev, E. Biyik, D. A. Lazar, W. Z. Wang, D. Sadigh, and R. Pedarsani, "Incentivizing routing choices for safe and efficient transportation in the face of the covid-19 pandemic," in *12th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPs)*, May 2021.
- [54] E. Biyik, D. A. Lazar, R. Pedarsani, and D. Sadigh, "Incentivizing efficient equilibria in traffic networks with mixed autonomy," *IEEE Transactions on Control of Network Systems (TCNS)*, 2021.
- [55] E. Biyik, F. Yao, Y. Chow, A. Haig, C.-w. Hsu, M. Ghavamzadeh, and C. Boutilier, "Preference elicitation with soft attributes in interactive recommendation," *arXiv preprint arXiv:2311.02085*, 2023.

- [56] E. Biyik and D. Sadigh, "Batch active preference-based learning of reward functions," in *Proceedings of the 2nd Conference on Robot Learning (CoRL)*, ser. Proceedings of Machine Learning Research, vol. 87, PMLR, 2018, pp. 519–528.
- [57] A. Kulesza, B. Taskar, et al., "Determinantal point processes for machine learning," *Foundations and Trends® in Machine Learning*, vol. 5, no. 2–3, pp. 123–286, 2012.
- [58] E. Biyik, K. Wang, N. Anari, and D. Sadigh, "Batch active learning using determinantal point processes," *arXiv preprint arXiv:1906.07975*, 2019.
- [59] E. Biyik, N. Anari, and D. Sadigh, "Batch active learning of reward functions from human preferences," *ACM Transactions on Human-Robot Interaction (THRI)*, 2024.
- [60] Z. Zhao, P. Piech, and L. Xia, "Learning mixtures of plackett-luce models," in *International Conference on Machine Learning*, PMLR, 2016, pp. 2906–2914.
- [61] V. Myers, E. Biyik, N. Anari, and D. Sadigh, "Learning multimodal rewards from rankings," in *Proceedings of the 5th Conference on Robot Learning (CoRL)*, 2021.
- [62] C. Basu, E. Biyik, Z. He, M. Singhal, and D. Sadigh, "Active learning of reward dynamics from hierarchical queries," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Nov. 2019.
- [63] N. Wilde, E. Biyik, D. Sadigh, and S. L. Smith, "Learning reward functions from scale feedback," in *Proceedings of the 5th Conference on Robot Learning (CoRL)*, 2021.
- [64] Z. Yang, M. Jun, J. Tien, S. J. Russell, A. Dragan, and E. Biyik, "Trajectory improvement and reward learning from comparative language feedback," in *Conference on Robot Learning (CoRL)*, 2024.
- [65] S. A. Sontakke, J. Zhang, S. M. R. Arnold, K. Pertsch, E. Biyik, D. Sadigh, C. Finn, and L. Itti, "Roboclip: One demonstration is enough to learn robot policies," in *Conference on Neural Information Processing Systems (NeurIPS)*, 2023.
- [66] Y. Wang, Z. Sun, J. Zhang, Z. Xian, E. Biyik, D. Held, and Z. Erickson, "Rl-vlm-f: Reinforcement learning from vision language foundation model feedback," in *International Conference on Machine Learning (ICML)*, 2024.
- [67] Y. Ling, K. Owalekar, O. Adesanya, E. Biyik, and D. Seita, "Impact: Intelligent motion planning with acceptable contact trajectories via vision-language models," *arXiv preprint arXiv:2503.10110*, 2025.
- [68] J. Zhang, Y. Luo, A. Anwar, S. A. Sontakke, J. J. Lim, J. Thomason, E. Biyik, and J. Zhang, "Rewind: Language-guided rewards teach robot policies without new demonstrations," in *Conference on Robot Learning (CoRL)*, 2025.
- [69] A. O'Neill, A. Rehman, A. Maddukuri, A. Gupta, A. Padalkar, A. Lee, A. Pooley, A. Gupta, A. Mandlekar, A. Jain, et al., "Open x-embodiment: Robotic learning datasets and rt-x models: Open x-embodiment collaboration 0," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2024, pp. 6892–6903.
- [70] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, et al., "Dinov2: Learning robust visual features without supervision," *Transactions on Machine Learning Research Journal*, pp. 1–31, 2024.
- [71] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Association for Computational Linguistics, 2019, p. 3982.
- [72] T. Yu, D. Quillen, Z. He, R. Julian, K. Hausman, C. Finn, and S. Levine, "Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning," in *Conference on robot learning*, PMLR, 2020, pp. 1094–1100.
- [73] D. J. Foster, A. Block, and D. Misra, "Is behavior cloning all you need? understanding horizon in imitation learning," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.
- [74] X. Liu, T. Yoneda, R. Stevens, M. Walter, and Y. Chen, "Blending imitation and reinforcement learning for robust policy improvement," in *The Twelfth International Conference on Learning Representations*, 2024.
- [75] T. Hester, M. Vecerik, O. Pietquin, M. Lanctot, T. Schaul, B. Piot, D. Horgan, J. Quan, A. Sendonaris, I. Osband, et al., "Deep q-learning from demonstrations," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.
- [76] H. Hu, S. Mirchandani, and D. Sadigh, "Imitation bootstrapped reinforcement learning," *Proceedings of Robotics: Science and Systems (RSS)*, 2024.
- [77] M. Albaba, S. Christen, T. Langarek, C. Gebhardt, O. Hilliges, and M. J. Black, "Rile: Reinforced imitation learning," *arXiv preprint arXiv:2406.08472*, 2024.
- [78] W. Sun, J. A. Bagnell, and B. Boots, "Truncated horizon policy search: Combining reinforcement learning & imitation learning," in *International Conference on Learning Representations*, 2018.
- [79] S. Noh, S. Kim, and I. Jang, "Efficient fine-tuning of behavior cloned policies with reinforcement learning from limited demonstrations," in *NeurIPS 2024 Workshop on Fine-Tuning in Modern Machine Learning: Principles and Scalability*.
- [80] Z.-H. Yin, W. Ye, Q. Chen, and Y. Gao, "Planning for sample efficient imitation learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 2577–2589, 2022.
- [81] M. Hong, A. Liang, K. Kim, H. Rajaprakash, J. Thomason, E. Biyik, and J. Zhang, "Hand me the data: Fast robot adaptation via hand path retrieval," *arXiv preprint arXiv:2505.20455*, 2025.
- [82] M. Müller, *Fundamentals of music processing: Using Python and Jupyter notebooks*. Springer, 2021, vol. 2.
- [83] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, "Lora: Low-rank adaptation of large language models," *arXiv preprint arXiv:2106.09685*, 2021.
- [84] O. Mees, L. Hermann, E. Rosete-Beas, and W. Burgard, "Calvin: A benchmark for language-conditioned policy learning for long-horizon robot manipulation tasks," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 7327–7334, 2022.
- [85] V. Myers, E. Biyik, and D. Sadigh, "Active reward learning from online preferences," in *International Conference on Robotics and Automation (ICRA)*, May 2023.
- [86] P. Viappiani and C. Boutilier, "Optimal bayesian recommendation sets and myopically optimal choice query sets," *Advances in neural information processing systems*, vol. 23, 2010.
- [87] R. Cohn, E. Durfee, and S. Singh, "Comparing action-query strategies in semi-autonomous agents," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 25, 2011, pp. 1102–1107.
- [88] Y. Korkmaz and E. Biyik, "Mile: Model-based intervention learning," in *International Conference on Robotics and Automation (ICRA)*, 2025.
- [89] A. Liang, J. Thomason, and E. Biyik, "Visarl: Visual reinforcement learning guided by human saliency," in *International Conference on Intelligent Robots and Systems (IROS)*, 2024.
- [90] A. Banayeezade, F. Bahrani, Y. Zhou, and E. Biyik, "Gabril: Gaze-based regularization for mitigating causal confusion in imitation learning," in *International Conference on Intelligent Robots and Systems (IROS)*, 2025.
- [91] L.-H. Lin, Y. Cui, Y. Hao, F. Xia, and D. Sadigh, "Gesture-informed robot assistance via foundation models," in *7th Annual Conference on Robot Learning*, 2023.
- [92] M. Stiber, R. Taylor, and C.-M. Huang, "Robot error awareness through human reactions: Implementation, evaluation, and recommendations," *arXiv preprint arXiv:2501.05723*, 2025.