Interactive Imitation Learning in Robotics: A Survey

Other titles in Foundations and Trends® in Robotics

A Survey on the Integration of Machine Learning with Sampling-based Motion Planning
Troy McMahon, Aravind Sivaramakrishnan, Edgar Granados and Kostas E. Bekris
ISBN: 978-1-63828-134-4

Adoption of Robots for Disasters: Lessons from the Response to COVID-19
Robin R. Murphy, Vignesh B.M. Gandudi, Justin Adams, Angela Clendenin and Jason Moats
ISBN: 978-1-68083-862-6

Cybersecurity in Robotics: Challenges, Quantitative Modeling, and Practice
Quanyan Zhu, Stefan Rass, Bernhard Dieber and Víctor Mayoral Vilches
ISBN: 978-1-68083-860-2

A Roadmap for US Robotics – From Internet to Robotics 2020 Edition
Henrik Christensen, Nancy Amato, Holly Yanco, Maja Mataric, Howie Choset, Ann Drobnis, Ken Goldberg, Jessy Grizzle, Gregory Hager, John Hollerbach, Seth Hutchinson, Venkat Krovi, Daniel Lee, Bill Smart, Jeff Trinkle and Gaurav Sukhatme
ISBN: 978-1-68083-858-9

The State of Industrial Robotics: Emerging Technologies, Challenges, and Key Research Directions
Lindsay Sanneman, Christopher Fourie and Julie A. Shah
ISBN: 978-1-68083-800-8

Interactive Imitation Learning in Robotics: A Survey

Carlos Celemin
Delft University of Technology
c.e.celeminpaez@tudelft.nl

Rodrigo Pérez-Dattari
Delft University of Technology
r.j.perezdattari@tudelft.nl

Eugenio Chisari
University of Freiburg
chisari@cs.uni-freiburg.de

Giovanni Franzese
Delft University of Technology
g.franzese@tudelft.nl

Leandro de Souza Rosa
Delft University of Technology
l.desouzarosa@tudelft.nl

Ravi Prakash
Delft University of Technology
r.prakash-1@tudelft.nl

Zlatan Ajanović
Delft University of Technology
z.ajanovic@tudelft.nl

Marta Ferraz
Delft University of Technology
m.ferraz@tudelft.nl

Abhinav Valada
University of Freiburg
valada@cs.uni-freiburg.de

Jens Kober
Delft University of Technology
j.kober@tudelft.nl

the essence of knowledge
Boston — Delft

Foundations and Trends® in Robotics

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

C. Celemin *et al.*. *Interactive Imitation Learning in Robotics: A Survey*. Foundations and Trends® in Robotics, vol. 10, no. 1-2, pp. 1–197, 2022.

ISBN: 978-1-63828-127-6

Foundations and Trends® in Robotics
Volume 10, Issue 1-2, 2022

Editorial Scope

Topics

Foundations and Trends® in Robotics publishes survey and tutorial articles in the following topics:

- Mathematical modelling
- Kinematics
- Dynamics
- Estimation Methods
- Robot Control
- Planning
- Artificial Intelligence in Robotics
- Software Systems and Architectures
- Mechanisms and Actuators
- Sensors and Estimation
- Planning and Control
- Human-Robot Interaction
- Industrial Robotics
- Service Robotics

Information for Librarians

Foundations and Trends® in Robotics, 2022, Volume 10, 4 issues. ISSN paper version 1935-8253. ISSN online version 1935-8261. Also available as a combined paper and online subscription.

Contents

Interactive Imitation Learning in Robotics: A Survey

Carlos Celemin[1*], Rodrigo Pérez-Dattari[1*], Eugenio Chisari[2*], Giovanni Franzese[1*], Leandro de Souza Rosa[1], Ravi Prakash[1], Zlatan Ajanović[1], Marta Ferraz[1], Abhinav Valada[2] and Jens Kober[1]

[1] *Department of Cognitive Robotics, Delft University of Technology, The Netherlands*

[2] *Robot Learning Lab, University of Freiburg, Germany*

ABSTRACT

Interactive Imitation Learning (IIL) is a branch of Imitation Learning (IL) where human feedback is provided intermittently during robot execution allowing an online improvement of the robot's behavior.

In recent years, IIL has increasingly started to carve out its own space as a promising data-driven alternative for solving complex robotic tasks. The advantages of IIL are twofold, 1) it is data-efficient, as the human feedback guides the robot directly towards an improved behavior (in contrast with Reinforcement Learning (RL), where behaviors must be discovered by trial and error), and 2) it is robust, as the distribution mismatch between the teacher and learner trajectories is minimized by providing feedback directly over the learner's trajectories (as opposed to offline IL methods such as Behavioral Cloning).

Nevertheless, despite the opportunities that IIL presents, its terminology, structure, and applicability are not clear

*These authors contributed equally to this work.

Carlos Celemin, Rodrigo Pérez-Dattari, Eugenio Chisari, Giovanni Franzese, Leandro de Souza Rosa, Ravi Prakash, Zlatan Ajanović, Marta Ferraz, Abhinav Valada and Jens Kober (2022), "Interactive Imitation Learning in Robotics: A Survey", Foundations and Trends® in Robotics: Vol. 10, No. 1-2, pp 1–197. DOI: 10.1561/2300000072.

nor unified in the literature, slowing down its development and, therefore, the research of innovative formulations and discoveries.

In this work, we attempt to facilitate research in IIL and lower entry barriers for new practitioners by providing a survey of the field that unifies and structures it. In addition, we aim to raise awareness of its potential, what has been accomplished and what are still open research questions.

We organize the most relevant works in IIL in terms of human-robot interaction (i.e., types of feedback), interfaces (i.e., means of providing feedback), learning (i.e., models learned from feedback and function approximators), user experience (i.e., human perception about the learning process), applications, and benchmarks. Furthermore, we analyze similarities and differences between IIL and RL, providing a discussion on how the concepts *offline*, *online*, *off-policy* and *on-policy* learning should be transferred to IIL from the RL literature.

We particularly focus on robotic applications in the real world and discuss their implications, limitations, and promising future areas of research.

1

Introduction

1.1 Motivation

Existing robotic technology is still mostly limited to being used by expert programmers who can adapt the systems to new required conditions, but not flexible and adaptable by non-expert workers or end-users. Imitation Learning (IL) has obtained considerable attention as a potential direction for enabling all kinds of users to easily program the behavior of robots or virtual agents. The teaching process takes place directly in the application context, in a natural way for humans, and does not require engineering effort to adapt the behavior for each different scenario.

In the case teachers (i.e., humans with knowledge about the task) are available and able to transfer their knowledge to the agent, it is preferred to program behaviors from recorded demonstrations rather than tackling the problem with other Machine Learning (ML) techniques such as Reinforcement Learning (RL), which involve additional design, infrastructure, safety, and data efficiency challenges (Sutton and Barto, 2018), and in many cases are not applicable to physical systems due to time and resource limitations.

When considering the advantages of programming robots in a natural way, like we humans do for teaching complex skills (e.g., requiring fast dynamics and dexterity) to others, the possibilities are not limited to recording demonstrations, for later fitting a policy model, as it is done in traditional IL methods (Argall *et al.*, 2009). In practice, an initial set of demonstrations or instructions tend to suffice to teach very simple and easy tasks from human to human, e.g., the instructions for opening a door, plugging a phone charger, or the user guide for most devices we use on a daily basis. Nevertheless, for complex skills such as playing a sport, a loop of interactions is required for learning, because then the teacher explains/shows the student what to do by directly correcting/evaluating its actions, improving its behavior over past mistakes and successes. Otherwise, considering and explaining all possible scenarios in advance would be intractable for both the teacher and the student.

This kind of teaching is based on different types of teaching feedback, like demonstrations, sporadic corrections, or evaluations (grading) with value judgments or rankings. As an example, when teaching a complex skill like playing tennis, various steps can be involved. The teacher shows full demonstrations of the stroke themselves to the learner. When the student tries to replicate the example, the teacher can show what a better execution would look like. After the student performs the stroke, the teacher could advise with voice instructions to slightly correct the angles, velocities, or forces of the movement. Moreover, the teacher can sporadically congratulate the student or make it clear that some decisions were not so good. This kind of interactive teaching approach seems to be, for humans, the most natural strategy for teaching to perform more complex skills; therefore, it is desirable to teach robots in the same way.

In recent years, the domains of robotics and ML have increasingly adopted and developed these interactive teaching strategies, as can be observed in Figure 1.1. In this work, Interactive Imitation Learning (IIL) refers to all the methods that include the teacher in the learning loop for training sequential decision-making systems. The objective of this work is to survey the literature on these methods and to present the most relevant observations in an organized structure.

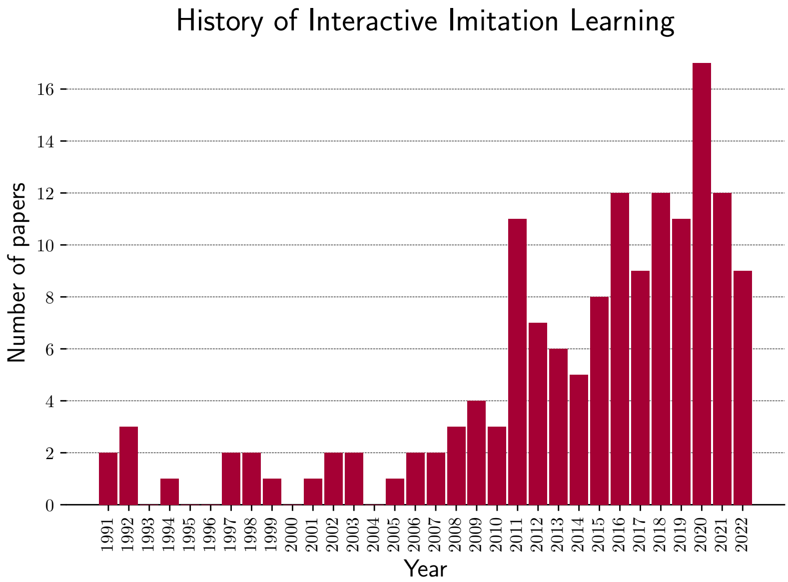

Figure 1.1: Histogram of IIL papers (from the group of works surveyed until the beginning of 2022) written per year.

The study of IIL methods has increased and the community has grown because these strategies introduce additional benefits with respect to learning paradigms such as traditional IL. Some of those advantages are:

- A more natural or intuitive teaching approach.
- Enabling users who are non-experts at demonstrating the task to teach successful policies.
- Obtaining richer datasets consisting of data from situations that are not faced when learning from full demonstrations, as the distribution of data collected is induced by the learner instead of the teacher, avoiding data mismatch issues (see Section 2.2.6).
- More flexibility to the teachers, who are not constrained to use only demonstrations for transferring their knowledge, but they

can use other kinds of feedback, like relative corrections, human reinforcements, or comparisons.

- Offers alternatives to solve the correspondence problem that exists between the space where teachers can give demonstrations and the space where the robot executes the actions.
- Some methods have more tolerance for the teacher's mistakes or provide a better approach to compensate for them.

Nonetheless, there are certain challenges that should be considered when a teacher is in the learning loop. Human teachers can be inconsistent and make mistakes, there is uncertainty in their input that tries to explain their intention, they need to learn to adapt to the changing behavior of the learning agent, and the learning process is open-ended (Dudley and Kristensson, 2018).

In this work, we review the context that defines the domain of IIL and how it relates to other known learning approaches. We highlight the most relevant aspects to be considered for teaching an agent interactively and organize the methods according to them. This study is based on grouping and surveying the most relevant established papers in the literature, along with more recent follow-up works that have shown promising contributions. All these papers were gathered in a set of works used as reference for organizing the different classifications proposed throughout the different sections. This set is also used for generating the tables in Sections 3 and 4, and the plot of Figure 1.1.

One of the reasons such organization of IIL methods does not exist so far is due to the varied terminology used by different authors to refer to some of these methods, which in many cases, only partially overlap. Below, we introduce most of the names and keywords used to refer to the approaches that are relevant in this work.

1.2 Terminology Unification

In the literature, there are many terms linked to ML approaches that enable teachers to interactively shape learning systems. As a consequence, many of them are used to describe similar learning problems,

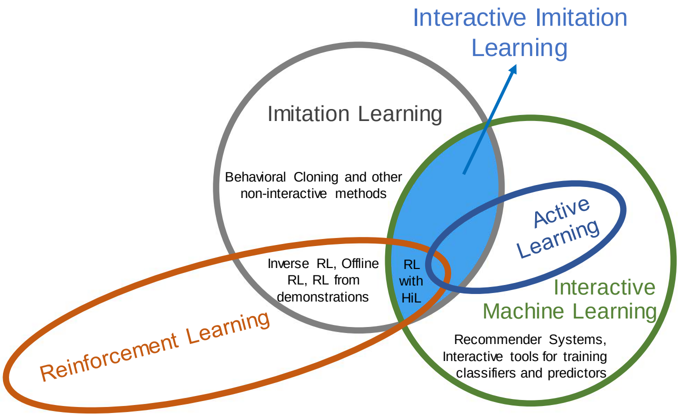

Figure 1.2: Relationship between different sets of learning paradigms related to the scope of this work. The intersection of IL with Interactive Machine Learning (IML)(blue area) is what defines the scope of this work, called here IIL

which makes it difficult for practitioners (especially beginners) to have a clear outlook of the field when studying the well-spread collection of related papers. In this section, we introduce some of those terms and discuss how they relate to each other, group them into sets that partially overlap or contain some others, and provide a definition of IIL. Based on this definition and structure, we set the bounds of the topic of interest of this work.

Figure 1.2 depicts with a Venn Diagram the relationship between all learning paradigms discussed below.

1.2.1 Imitation Learning

In the context of robotics, the terms Learning from Demonstration (LfD), Programming by Demonstrations or Programming by Doing (PbD), and IL are indistinctly used when referring to the paradigm of enabling robots to derive controllers from human demonstrations (Billard and Grollman, 2013). Originally, these terms have been used by multiple authors referring to learning approaches that derive policies from datasets of explicit teacher demonstrations of a task.

Some recent methods enable human teachers to train robots through evaluative feedback, like Learning from Critique (LfC), or Interactive Reinforcement Learning (Interactive RL), in which the teachers provide feedback that rates the desirability of the exhibited behavior during training time. Although these approaches do not fully fit the literal meaning of LfD or IL, some authors consider that evaluative feedback is just one of the demonstration modes a teacher could use within a learning process (Chernova and Thomaz, 2014), therefore they also can be considered part of the world of IL.

Since IL is used at different levels of robot control and similar problems, we can rephrase the definitions of LfD, PbD, and IL as *the set of ML methods that leverage teacher's input as the source of knowledge for training sequential decision-making systems*. Most of the time, the teacher is a human user, while in some cases it could be another decision-making agent (e.g., a computationally expensive policy like an Model Predictive Control (MPC) or a planner system), and it has an understanding of either what are the objectives of the task, what to do, how good an action/policy is, or how good is the policy with respect to others.

In other words, methods are not considered IL if they leverage the input of a teacher to train non-sequential decision-making systems, e.g. an image classifier (Fails and Olsen Jr, 2003).

In the last two decades, articles have been published reviewing varied perspectives of IL, proposing categorizations for organizing the types of methods, identifying the benefits and drawbacks of the most known approaches, listing the open challenges, and introducing and structuring the field of study (Billard *et al.*, 2008; Argall *et al.*, 2009; Billing and Hellström, 2010; Billard and Grollman, 2013; Chernova and Thomaz, 2014; Amershi *et al.*, 2014; Billard *et al.*, 2016; Hussein *et al.*, 2017; Lee, 2017; Calinon, 2018; Osa *et al.*, 2018; Li *et al.*, 2019a; Zhang *et al.*, 2019b; Ravichandar *et al.*, 2020).

1.2.2 Interactive Machine Learning

There exists a considerable amount of learning methods that leverage human teachers within the learning loop for training sequential and non-sequential decision-making systems. Through different types of interaction, they make use of the knowledge a human has about the process, without the need to hard-coding it. Therefore, these methods enable users who are not expert ML practitioners to train models according to their insights and intuition. The set of approaches that cover all the learning loop schemes involving humans transferring knowledge to the agent is known as IML (Amershi *et al.*, 2014; Fails and Olsen Jr, 2003; Ware *et al.*, 2001; Holzinger, 2016; Dudley and Kristensson, 2018; Jiang *et al.*, 2019).

Holzinger (2016) define *"IML-approaches as algorithms that can interact with both computational agents and human agents and can optimize their learning behavior through these interactions"*. Dudley and Kristensson (2018) explain the contrast between IML and classical ML as *"Interactive Machine Learning is distinct from classical machine learning in that human intelligence is applied through iterative teaching and model refinement in a relatively tight loop of set-and-check. In other words, the user provides additional information to the system to update the model, and the change in the model is reviewed against the user's design objective"*.

Some other authors refer to the same domain with a more explicit name like Human in the Loop Machine Learning (HIL-ML) (Xin *et al.*, 2018; Wu *et al.*, 2021). Other authors refer to it in a more general way, combining the term Artificial Intelligence (AI), e.g., with Human in the Loop Artificial Intelligence (HIL-AI) (Zanzotto, 2019), or Interactive Artificial Intelligence (IAI) (Wenskovitch and North, 2020). Human Centered Machine Learning (HCML) or Human Centered Artificial Intelligence (HCAI) is a larger domain that contains all the mentioned approaches with a human in the learning loop, additionally, it also includes the approaches based on ML/AI that have humans in the execution loop, i.e., systems that interact with humans as in ML/AI-based Human-Computer Interaction (HCI) or Human-Robot Interaction (HRI) systems.

Methods of IML serve a wide domain of applications, including classification, regression, image processing, information retrieval, anomaly detection, among other systems (Ware *et al.*, 2001; Fails and Olsen Jr, 2003; Amershi *et al.*, 2012; Ngo *et al.*, 2014; Amershi *et al.*, 2014; Dudley and Kristensson, 2018; Jiang *et al.*, 2019). It is important to clarify that although IML methods always include a human in the learning process, in some applications the human does not always perform as a *teacher*, but rather is a user about whom the system learns through the interactions without explicit signals, as it is the case for Recommender Systems (Burke, 2002; Bobadilla *et al.*, 2013; Beel *et al.*, 2016).

Active Learning is one of the most traditional approaches of IML, which consists of endowing the learner with capabilities for querying the teacher for more data in specific situations. The learner is able to choose from which data samples it learns, allowing it to learn with higher accuracy from fewer samples (Cohn *et al.*, 1996; Settles, 2009).

1.2.3 Interactive Imitation Learning

The set of IML covers a broad spectrum of problems it can be applied to, including sequential and non-sequential decision-making. IL is narrower and specific to sequential problems. Unlike IML, IL also involves methods that learn from teachers in a sequential manner, without the need for continuous interaction in the learning loop, as is the case of Behavioral Cloning (BC), Inverse Reinforcement Learning (IRL), offline RL, or RL from demonstrations, which learn from a set of demonstrations that have been recorded before the learning process starts.

Also known as direct IL, BC (Bain and Sammut, 1995) applies supervised learning to a set of previously recorded expert demonstrations, in order to obtain a model that imitates the demonstrations. In contrast, IRL is known as indirect IL because it uses recorded demonstrations to obtain an objective function or reward function that explains the goal of the task, so it can be used in an RL process for obtaining a policy that imitates the demonstrator (Ng and Russell, 2000; Zhifei and Joo, 2012). In offline RL the principles of classical online RL are extended to be applied over datasets of demonstrations, without collecting any new sample during training time (Levine *et al.*, 2020). We refer to RL from demonstrations to the domain of all methods of classical online RL that leverage recorded demonstrations to initialize the policy, or that keep that data in a buffer that is continuously used for updating the policy along with the new samples that are collected with the interactions (Kober and Peters, 2008; Hester *et al.*, 2018).

The previous methods are not interactive, even though they learn from data demonstrated by teachers. We hereby, take the term IIL that has been previously used in the literature and redefine it as the set of methods resulting from the intersection of the IL and IML sets. Therefore, we can say that *IIL methods involve the approaches that learn from the knowledge provided by a teacher in the learning loop of a sequential decision-making system*. Human teachers can transfer their knowledge to the learning agent through different modalities of interaction, and they are able to observe the effect of their feedback throughout the incremental learning process.

Methods of ML that actively choose or query training samples are known as Active Learning (Settles, 2009) methods, and they aim to increase the sampling efficiency of the learning process. It is a subset of IML that also overlaps with the IIL domain.

It is important to make a distinction between IIL, IML, and Interactive Learning Systems (ILS), which is also used in the literature and sometimes referred to as learning from interactions, or interactive learning. ILS are real/virtual entities that learn from the interaction with the world, a human, or another entity. This definition is complemented in Cuayáhuitl *et al.* (2013) with the description: *"A machine can therefore be said to learn from interactions in a particular class of tasks if its performance improves with the given interactions over time"*. The ILS that learn from the interaction with the world/environment enclose RL methods (Sutton and Barto, 2018), wherein the agent learns from its own experience and not from a teacher. The subset of ILS that learn from the interaction with other agents acting as teachers results in the same set of IIL methods, which are the focus of this work.

RL systems that obtain data from human teachers in the form of either demonstrations or evaluations (human reinforcements) during the learning process are known as Human in the Loop Reinforcement Learning (HIL-RL) and are also a type of IIL.

1.3 Others Surveys and Outline

In recent years, there has been an explosion in the adoption of IL methods. There exist a large body of surveys discussing IL from different points of view. In particular, Chernova and Thomaz (2014) provides a general overview of the methodology of learning from demonstration where different topics are analyzed, such as how the human teacher interacts with the robot to provide demonstrations, which modeling technique to choose (low/high level), how the human can refine an existing task and how to incorporate interactive and active learning components. Given the big spectrum of the paper of Chernova and Thomaz (2014), interactive methods are mentioned as one possible evolution of IL, but they are not the main focus of the work, and, therefore, not analyzed in depth.

A similar collection and analysis of the literature were conducted recently by Ravichandar *et al.* (2020). Here, topics such as non-expert robot programming, data efficiency, safe learning, and performance guarantees are discussed. The authors highlight the importance of learning from social cues, reasoning about the availability of human demonstrators, how to behave in their absence and how to ask for help. However, IIL is only marginally analyzed.

Similarly, Hussein *et al.* (2017) propose a survey on different learning methods for IL. The survey underlines how BC has limitations due to errors in the demonstration and poor generalization. As a possible solution, it is proposed to combine IL with RL, refine the policy with RL, or use active learning. However, marginal attention is given specifically to interactive methods.

In a recent survey, Osa *et al.* (2018) provide a structural analysis on IL, focusing on BC and IRL methods. The authors mention that incremental and interactive learning methods can be employed to alleviate the *covariate shift* problem (Section 2.2.6) that exist in BC methods. While they highlight the necessity of such methods from an algorithmic and mathematical perspective on machine learning, the authors do not provide an extensive treatment of the topic, as it is outside the scope of their work.

The topic of Human-Centered RL is investigated by (Li *et al.*, 2019a) as well as Zhang *et al.* (2019b), where human evaluative feedback is used to teach behaviors to learning agents. They divide the field into three categories: learning from human reward, from interpreted human reward, and from action-translated human reward. Although these works are surveying the concept of human feedback from a RL perspective, a broader discussion of other IIL methods is not covered.

In our work, we provide a survey of the Interactive Imitation Learning literature, ranging from seminal early work to the most recent advances. We investigate the role of IIL in the broader picture of sequential decision-making problems, with a focus on robotics applications. Besides providing an organized view of the state-of-the-art of the field, we aim to distill the most important takeaways and contribute a useful perspective on the topic. Our goal is for this manuscript to be a helpful reference for future work as well as a starting point for newcomers to the field. Our discussion spans multiple dimensions, ranging from the type of feedback a human teacher can provide to the agents, to the models that are learned through this interaction, to the existing benchmarks and applications proposed in recent years. In particular, we structure the analysis over multiple sections as follows:

- Section 2 provides an overview of the sequential decision-making problem and its different formulations, formalizes the IIL problem and defines core concepts such as Feedback and Covariate Shift.

- Section 3 discusses the different modalities of feedback that a human teacher can provide to the robot, ranging from evaluative to preference to corrective feedback or interventions. We examine their strengths and weaknesses, with a focus on the trade-off between richness of information and human effort required.

- Section 4 considers the various types of models that the robot is able to learn from the provided feedback, including policies, transition models, and objective functions. We discuss how certain models are best learned by specific types of feedback, and how they are used to achieve the main objective of solving sequential decision problems.

- Section 5 reviews auxiliary models that the robot could learn in addition to the main objective, such as uncertainty and risk estimation models, environment dynamics, task features and models of the human teacher. We analyze the advantages that such models provide and the settings in which they can be adopted.

- Section 6 discusses the different types of function approximation and model representation strategies common in the literature, including motion-conditioned models and deep neural networks. We consider their advantages and disadvantages and provide recommendations on their usage.
- Section 7 provides a comparison between on-policy and off-policy methods with a focus on the IIL setting.
- Section 8 analyzes the special case of IIL methods used in glsrl framework, called RL with Human in the Loop.
- Section 9 presents an overview of the interfaces used for enabling the communication between the robot/computer and the teacher, examining their role and importance in the learning pipeline. They range from physical contact with the robot embodiment to external devices such as remote controllers to contact-free approaches such as video and voice.
- Section 10 provides an overview of the human factors to consider in IIL, such as available human-robot interfaces, user experience, and performance metrics, as well as guidelines on how to design user studies in IIL.
- Section 11 surveys the principal benchmarks and datasets used in the literature to evaluate the proposed methods as well as the different fields of application of these algorithms, such as assistive, household, medical or industrial robots;
- Section 12 provides a discussion of the current challenges and opportunities in the field of IIL, as well as directions for future work.
- Section 13 completes the survey with a summary of the main concepts discussed as well as the most relevant takeaways and contributions to the field.

References

Abdel-Malek, K., J. Yang, W. Yu, and J. Duncan. (2005). "Human performance measures: mathematics". *Department of Mechanical Engineering The University of Iowa, Technical report*: 1–27.

Ablett, T., F. Marić, and J. Kelly. (2020). "Fighting Failures with FIRE: Failure Identification to Reduce Expert Burden in Intervention-Based Learning". *arXiv preprint arXiv:2007.00245*.

Akrour, R., M. Schoenauer, and M. Sebag. (2012). "APRIL: Active Preference Learning-Based Reinforcement Learning". In: *Machine Learning and Knowledge Discovery in Databases*. Berlin, Heidelberg: Springer Berlin Heidelberg. 116–131.

Akrour, R., M. Schoenauer, and M. Sebag. (2011). "Preference-Based Policy Learning". In: *Proceedings of the 2011 European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part I*. *ECML PKDD'11*. Athens, Greece: Springer-Verlag. 12–27.

Akrour, R., M. Schoenauer, M. Sebag, and J.-C. Souplet. (2014). "Programming by Feedback". In: *International Conference on Machine Learning*. *JMLR Proceedings*. No. 32. Pékin, China: JMLR.org. 1503–1511. URL: https://hal.inria.fr/hal-00980839.

Alshiekh, M., R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu. (2018). "Safe Reinforcement Learning via Shielding". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1.

Aly, A., S. Griffiths, and F. Stramandinoli. (2017). "Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics". *Cognitive Systems Research*. 43: 313–323. DOI: https://doi.org/10.1016/j.cogsys.2016.06.002.

Amershi, S., M. Cakmak, W. B. Knox, and T. Kulesza. (2014). "Power to the people: The role of humans in interactive machine learning". *AI Magazine*. 35(4): 105–120.

Amershi, S., J. Fogarty, and D. Weld. (2012). "Regroup: Interactive machine learning for on-demand group creation in social networks". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 21–30.

Amodei, D., C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané. (2016). "Concrete Problems in AI Safety". DOI: 10.48550/arXiv.1606.06565.

Arakawa, R., S. Kobayashi, Y. Unno, Y. Tsuboi, and S.-i. Maeda. (2018). "DQN-TAMER: Human-in-the-Loop Reinforcement Learning with Intractable Feedback". DOI: 10.48550/ARXIV.1810.11748.

Argall, B. D. (2009). "Learning mobile robot motion control from demonstration and corrective feedback". *PhD thesis*. Carnegie Mellon University.

Argall, B. D., S. Chernova, M. Veloso, and B. Browning. (2009). "A survey of robot learning from demonstration". *Robotics and autonomous systems*. 57(5): 469–483.

Argall, B. D., B. Browning, and M. Veloso. (2008). "Learning robot motion control with demonstration and advice-operators". In: *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 399–404. DOI: 10.1109/IROS.2008.4651020.

Argall, B. D., B. Browning, and M. M. Veloso. (2011a). "Teacher feedback to scaffold and refine demonstrated motion primitives on a mobile robot". *Robotics and Autonomous Systems*. 59(3): 243–255. DOI: https://doi.org/10.1016/j.robot.2010.11.004.

Argall, B. D., E. L. Sauser, and A. G. Billard. (2011b). "Tactile Guidance for Policy Adaptation". *Foundations and Trends® in Robotics*. 1(2): 79–133. DOI: 10.1561/2300000012.

Arumugam, D., J. K. Lee, S. Saskin, and M. L. Littman. (2019). "Deep Reinforcement Learning from Policy-Dependent Human Feedback". DOI: 10.48550/ARXIV.1902.04257.

Arzate Cruz, C. and T. Igarashi. (2020). "A Survey on Interactive Reinforcement Learning: Design Principles and Open Challenges". In: Proceedings of the 2020 ACM Designing Interactive Systems Conference. 1195–1209.

Bain, M. and C. Sammut. (1995). "A Framework for Behavioural Cloning." In: *Machine Intelligence 15*. 103–129.

Bajcsy, A., D. P. Losey, M. K. O'Malley, and A. D. Dragan. (2018). "Learning from physical human corrections, one feature at a time". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 141–149.

Bajcsy, A., D. P. Losey, M. K. O'Malley, and A. D. Dragan. (2017). "Learning robot objectives from physical human interaction". In: *Conference on Robot Learning*. PMLR. 217–226.

Balakrishna, A., B. Thananjeyan, J. Lee, F. Li, A. Zahed, J. E. Gonzalez, and K. Goldberg. (2020). "On-policy robot imitation learning from a converging supervisor". In: *Conference on Robot Learning*. PMLR. 24–41.

Beel, J., B. Gipp, S. Langer, and C. Breitinger. (2016). "Paper recommender systems: a literature survey". *International Journal on Digital Libraries*. 17(4): 305–338.

Behnke, S. (2006). "Robot competitions-ideal benchmarks for robotics research". In: *Proc. of IROS-2006 Workshop on Benchmarks in Robotics Research*. Institute of Electrical and Electronics Engineers (IEEE).

Bellman, R. (1957). "A Markovian decision process". *Journal of Mathematics and Mechanics*: 679–684.

Ben Amor, H., E. Berger, D. Vogt, and B. Jung. (2009). "Kinesthetic Bootstrapping: Teaching Motor Skills to Humanoid Robots through Physical Interaction". In: *KI 2009: Advances in Artificial Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg. 492–499.

Billard, A., S. Calinon, R. Dillmann, and S. Schaal. (2008). "Survey: Robot programming by demonstration". *Handbook of robotics*. 59(BOOK_CHAP).

Billard, A. and D. Grollman. (2013). "Robot learning by demonstration". *Scholarpedia*. 8(12): 3824.

Billard, A. G., S. Calinon, and R. Dillmann. (2016). "Learning from humans". In: *Springer handbook of robotics*. Springer. 1995–2014.

Billing, E. A. and T. Hellström. (2010). "A formalism for learning from demonstration". *Paladyn, Journal of Behavioral Robotics*. 1(1): 1–13.

Bishop, C. M. (2006). "Pattern recognition". *Machine learning*. 128(9).

Biyik, E., M. Palan, N. C. Landolfi, D. P. Losey, and D. Sadigh. (2020). "Asking Easy Questions: A User-Friendly Approach to Active Reward Learning". In: *Proceedings of the Conference on Robot Learning*. Vol. 100. *Proceedings of Machine Learning Research*. PMLR. 1177–1190. URL: https://proceedings.mlr.press/v100/b-iy-ik20a.html.

Biyik, E. and D. Sadigh. (2018). "Batch Active Preference-Based Learning of Reward Functions". In: *Proceedings of The 2nd Conference on Robot Learning*. Vol. 87. *Proceedings of Machine Learning Research*. PMLR. 519–528. URL: https://proceedings.mlr.press/v87/biyik18a.html.

Bıyık, E., N. Huynh, M. J. Kochenderfer, and D. Sadigh. (2020). "Active Preference-Based Gaussian Process Regression for Reward Learning". DOI: 10.48550/ARXIV.2005.02575.

Blukis, V., N. Brukhim, A. Bennett, R. A. Knepper, and Y. Artzi. (2018). "Following high-level navigation instructions on a simulated quadcopter with imitation learning". *arXiv preprint arXiv:1806.00047*.

Blumberg, B., M. Downie, Y. Ivanov, M. Berlin, M. P. Johnson, and B. Tomlinson. (2002). "Integrated Learning for Interactive Synthetic Characters". In: *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*. *SIGGRAPH '02*. San Antonio, Texas: Association for Computing Machinery. 417–426. DOI: 10.1145/566570.566597.

Bobadilla, J., F. Ortega, A. Hernando, and A. Gutiérrez. (2013). "Recommender systems survey". *Knowledge-based systems*. 46: 109–132.

Bobu, A. and A. Peng. (2022). "Aligning Robot Representations with Humans". DOI: 10.48550/ARXIV.2205.07882.

Bobu, A., M. Wiggert, C. Tomlin, and A. D. Dragan. (2021). "Feature Expansive Reward Learning: Rethinking Human Input". In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction. HRI '21*. Boulder, CO, USA: Association for Computing Machinery. 216–224. DOI: 10.1145/3434073.3444667.

Bobu, A., M. Wiggert, C. Tomlin, and A. D. Dragan. (2022). "Inducing Structure in Reward Learning by Learning Features". *The International Journal of Robotics Research*. 0(0): 02783649221078031. DOI: 10.1177/02783649221078031.

Böhmer, W., J. T. Springenberg, J. Boedecker, M. Riedmiller, and K. Obermayer. (2015). "Autonomous learning of state representations for control: An emerging field aims to autonomously learn state representations for reinforcement learning agents from their real-world sensor observations". *KI-Künstliche Intelligenz*. 29(4): 353–362.

Bootsma, B., G. Franzese, and J. Kober. (2021). "Interactive learning of sensor policy fusion". In: *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*. IEEE. 665–670.

Bouthillier, X., P. Delaunay, M. Bronzi, A. Trofimov, B. Nichyporuk, J. Szeto, N. Mohammadi Sepahvand, E. Raff, K. Madan, V. Voleti, S. Ebrahimi Kahou, V. Michalski, T. Arbel, C. Pal, G. Varoquaux, and P. Vincent. (2021). "Accounting for Variance in Machine Learning Benchmarks". In: *Proceedings of Machine Learning and Systems*. Vol. 3. 747–769. URL: https://proceedings.mlsys.org/paper/2021/file/cfecdb276f634854f3ef915e2e980c31-Paper.pdf.

Brockman, G., V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. (2016). "OpenAI Gym". DOI: 10.48550/ARXIV.1606.01540.

Brown, D. S., Y. Cui, and S. Niekum. (2018). "Risk-aware active inverse reinforcement learning". In: *Conference on Robot Learning*. PMLR. 362–372.

Brown, D. S., R. Coleman, R. Srinivasan, and S. Niekum. (2020). "Safe Imitation Learning via Fast Bayesian Reward Inference from Preferences". In: *Proceedings of the 37th International Conference on Machine Learning. ICML'20*. JMLR.org.

Burke, R. (2002). "Hybrid recommender systems: Survey and experiments". *User modeling and user-adapted interaction*. 12(4): 331–370.

Cakmak, M. and A. L. Thomaz. (2012). "Designing robot learners that ask good questions". In: *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM. 17–24.

Calinon, S. (2018). "Learning from demonstration (programming by demonstration)". *Encyclopedia of Robotics*: 1–8.

Canal, G., G. Alenyà, and C. Torras. (2016). "Personalization Framework for Adaptive Robotic Feeding Assistance". In: *Social Robotics*. Cham: Springer International Publishing. 22–31.

Canal, G., E. Pignat, G. Alenyà, S. Calinon, and C. Torras. (2018). "Joining high-level symbolic planning with low-level motion primitives in adaptive HRI: application to dressing assistance". In: *IEEE International Conference on Robotics and Automation (ICRA)*. 3273–3278. DOI: 10.1109/ICRA.2018.8460606.

Canal, G., C. Torras, and G. Alenyà. (2021). "Are Preferences Useful for Better Assistance?: A Physically Assistive Robotics User Study". *ACM Transactions on Human-Robot Interaction (THRI)*. 10(4): 1–19. DOI: 10.1145/3472208.

Cederborg, T., I. Grover, C. L. Isbell Jr, and A. L. Thomaz. (2015). "Policy Shaping with Human Teachers." In: *IJCAI*. 3366–3372.

Celemin, C., G. Maeda, J. Ruiz-del-Solar, J. Peters, and J. Kober. (2019a). "Reinforcement learning of motor skills using policy search and human corrective advice". *The International Journal of Robotics Research*. 38(14): 1560–1580.

Celemin, C. and J. Ruiz-del-Solar. (2015). "COACH: Learning continuous actions from COrrective Advice Communicated by Humans". In: *2015 International Conference on Advanced Robotics (ICAR)*. 581–586. DOI: 10.1109/ICAR.2015.7251514.

Celemin, C. and J. Ruiz-del-Solar. (2019). "An interactive framework for learning continuous actions policies based on corrective feedback". *Journal of Intelligent & Robotic Systems*. 95(1): 77–97. DOI: 10.1007/s10846-018-0839-z.

Celemin, C., J. Ruiz-del-Solar, and J. Kober. (2019b). "A fast hybrid reinforcement learning framework with human corrective feedback". *Autonomous Robots*. 43(5): 1173–1186.

Chang, K.-W., A. Krishnamurthy, A. Agarwal, H. Daumé III, and J. Langford. (2015). "Learning to search better than your teacher". In: *International Conference on Machine Learning*. PMLR. 2058–2066.

Chatzimparmpas, A., R. M. Martins, I. Jusufi, and A. Kerren. (2020). "A survey of surveys on the use of visualization for interpreting machine learning models". *Information Visualization*. 19(3): 207–233.

Cheng, C.-A., X. Yan, N. Wagener, and B. Boots. (2018). "Fast Policy Learning through Imitation and Reinforcement". DOI: 10.48550/ARXIV.1805.10413.

Chernova, S. and A. L. Thomaz. (2014). "Robot learning from human teachers". *Synthesis Lectures on Artificial Intelligence and Machine Learning*. 8(3): 1–121.

Chernova, S. and M. Veloso. (2009). "Interactive policy learning through confidence-based autonomy". *Journal of Artificial Intelligence Research*. 34(1): 1.

Chisari, E., T. Welschehold, J. Boedecker, W. Burgard, and A. Valada. (2022). "Correct me if i am wrong: Interactive learning for robotic manipulation". *IEEE Robotics and Automation Letters*. 7(2): 3695–3702.

Christiano, P. F., J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei. (2017). "Deep Reinforcement Learning from Human Preferences". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper/2017/file/d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf.

Chu, V., B. Akgun, and A. L. Thomaz. (2016). "Learning haptic affordances from demonstration and human-guided exploration". In: *2016 IEEE Haptics Symposium (HAPTICS)*. 119–125. DOI: 10.1109/HAPTICS.2016.7463165.

Chu, V. and A. L. Thomaz. (2015). "Exploring Affordances Using Human-Guidance and Self-Exploration". In: *2015 AAAI Fall Symposium Series*.

Clouse, J. A. and P. E. Utgoff. (1992). "A Teaching Method for Reinforcement Learning". In: *Machine Learning Proceedings 1992*. Elsevier. 92–101.

Cohn, D. A., Z. Ghahramani, and M. I. Jordan. (1996). "Active learning with statistical models". *Journal of artificial intelligence research*. 4: 129–145.

Corrigan, L. J., C. Peters, D. Küster, and G. Castellano. (2016). "Engagement Perception and Generation for Social Robots and Virtual Agents". In: *Toward Robotic Socially Believable Behaving Systems - Volume I : Modeling Emotions*. Cham: Springer International Publishing. 29–51. DOI: 10.1007/978-3-319-31056-5_4.

Coumans, E. (2015). "Bullet Physics Simulation". In: *ACM SIGGRAPH 2015 Courses*. *SIGGRAPH '15*. Los Angeles, California: Association for Computing Machinery. DOI: 10.1145/2776880.2792704.

Cronrath, C., E. Jorge, J. Moberg, M. Jirstrand, and B. Lennartson. (2018). "BAgger: A Bayesian algorithm for safe and query-efficient imitation learning". In: *Machine Learning in Robot Motion Planning–IROS 2018 Workshop*.

Cruz, F., G. I. Parisi, and S. Wermter. (2018). "Multi-modal Feedback for Affordance-driven Interactive Reinforcement Learning". In: *2018 International Joint Conference on Neural Networks (IJCNN)*. 1–8. DOI: 10.1109/IJCNN.2018.8489237.

Cruz, F., J. Twiefel, S. Magg, C. Weber, and S. Wermter. (2015). "Interactive reinforcement learning through speech guidance in a domestic scenario". In: *2015 International Joint Conference on Neural Networks (IJCNN)*. 1–8. DOI: 10.1109/IJCNN.2015.7280477.

Cuayáhuitl, H., M. van Otterlo, N. Dethlefs, and L. Frommberger. (2013). "Machine learning for interactive systems and robots: a brief introduction". In: *Proceedings of the 2nd Workshop on Machine Learning for Interactive Systems: Bridging the Gap Between Perception, Action and Communication*. ACM. 19–28.

Cui, Y., D. Isele, S. Niekum, and K. Fujimura. (2019). "Uncertainty-aware data aggregation for deep imitation learning". In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE. 761–767.

Cui, Y., P. Koppol, H. Admoni, S. Niekum, R. Simmons, A. Steinfeld, and T. Fitzgerald. (2021). "Understanding the Relationship between Interactions and Outcomes in Human-in-the-Loop Machine Learning". In: *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Montreal, QC, Canada*. Vol. 10.

Cui, Y. and S. Niekum. (2018). "Active Reward Learning from Critiques". In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 6907–6914. DOI: 10.1109/ICRA.2018.8460854.

Curnow, H. J. and B. A. Wichmann. (1976). "A synthetic benchmark". *The Computer Journal*. 19(1): 43–49. DOI: 10.1093/comjnl/19.1.43.

CWI, I. and G. Amsterdam. (1997). "Cellular encoding for interactive evolutionary robotics". In: *Fourth European conference on artificial life*. Vol. 4. MIT Press. 368.

Daniel, C., O. Kroemer, M. Viering, J. Metz, and J. Peters. (2015). "Active Reward Learning with a Novel Acquisition Function". *Autonomous Robots*. 39(3): 389–405. DOI: 10.1007/s10514-015-9454-z.

Dasari, S., F. Ebert, S. Tian, S. Nair, B. Bucher, K. Schmeckpeper, S. Singh, S. Levine, and C. Finn. (2020). "RoboNet: Large-Scale Multi-Robot Learning". In: *Proceedings of the Conference on Robot Learning*. Vol. 100. *Proceedings of Machine Learning Research*. PMLR. 885–897. URL: https://proceedings.mlr.press/v100/dasari20a.html.

Degris, T., M. White, and R. Sutton. (2012). "Off-Policy Actor-Critic". In: *International Conference on Machine Learning*.

Deisenroth, M. P., G. Neumann, J. Peters, *et al.* (2013). "A survey on policy search for robotics". *Foundations and Trends® in Robotics*. 2(1–2): 1–142. URL: http://dx.doi.org/10.1561/2300000021.

Della Santina, C., C. Piazza, G. Grioli, M. G. Catalano, and A. Bicchi. (2018). "Toward dexterous manipulation with augmented adaptive synergies: The pisa/iit softhand 2". *IEEE Transactions on Robotics*. 34(5): 1141–1156.

DelPreto, J., J. I. Lipton, L. Sanneman, A. J. Fay, C. Fourie, C. Choi, and D. Rus. (2020). "Helping robots learn: a human-robot master-apprentice model using demonstrations via virtual reality teleoperation". In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 10226–10233.

Dosovitskiy, A., G. Ros, F. Codevilla, A. Lopez, and V. Koltun. (2017). "CARLA: An open urban driving simulator". In: *Conference on robot learning*. PMLR. 1–16.

Dudley, J. J. and P. O. Kristensson. (2018). "A Review of User Interface Design for Interactive Machine Learning". *ACM Trans. Interact. Intell. Syst.* 8(2). DOI: 10.1145/3185517.

Ewerton, M., G. Maeda, G. Kollegger, J. Wiemeyer, and J. Peters. (2016). "Incremental imitation learning of context-dependent motor skills". In: *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. 351–358. DOI: 10.1109/HUMANOIDS.2016.7803300.

Fails, J. A. and D. R. Olsen Jr. (2003). "Interactive machine learning". In: *Proceedings of the 8th international conference on Intelligent user interfaces*. ACM. 39–45.

Fan, L., Y. Zhu, J. Zhu, Z. Liu, O. Zeng, A. Gupta, J. Creus-Costa, S. Savarese, and L. Fei-Fei. (2018). "SURREAL: Open-Source Reinforcement Learning Framework and Robot Manipulation Benchmark". In: *Proceedings of The 2nd Conference on Robot Learning*. Vol. 87. *Proceedings of Machine Learning Research*. PMLR. 767–782. URL: https://proceedings.mlr.press/v87/fan18a.html.

Ferraz, M., E. Ferreira, E. d. Exter, F. v. d. Hulst, H. Rovina, W. Carey, J. Grenouilleau, and T. Krueger. (2019). "Multisensory real-time space telerobotics". In: *Intelligent Computing-Proceedings of the Computing Conference*. Springer. 275–298.

Finn, C., S. Levine, and P. Abbeel. (2016). "Guided Cost Learning: Deep Inverse Optimal Control via Policy Optimization". In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. ICML'16*. New York, NY, USA: JMLR.org. 49–58.

Fitzgerald, T., K. Bullard, A. Thomaz, and A. Goel. (2016). “Situated mapping for transfer learning”. In: *Fourth Annual Conference on Advances in Cognitive Systems*.

Fitzgerald, T., A. Goel, and A. Thomaz. (2018). “Human-Guided Object Mapping for Task Transfer”. *J. Hum.-Robot Interact.* 7(2). DOI: 10.1145/3277905.

Fleming, P. J. and J. J. Wallace. (1986). “How Not to Lie with Statistics: The Correct Way to Summarize Benchmark Results”. *Commun. ACM*. 29(3): 218–221. DOI: 10.1145/5666.5673.

Franzese, G., C. Celemin, and J. Kober. (2021a). “Learning Interactively to Resolve Ambiguity in Reference Frame Selection”. In: *Proceedings of the 2020 Conference on Robot Learning*. Vol. 155. *Proceedings of Machine Learning Research*. PMLR. 1298–1311. URL: https://proceedings.mlr.press/v155/franzese21a.html.

Franzese, G., A. Mészáros, L. Peternel, and J. Kober. (2021b). “ILoSA: Interactive Learning of Stiffness and Attractors”: 7778–7785. DOI: 10.1109/IROS51168.2021.9636710.

Fürnkranz, J., E. Hüllermeier, W. Cheng, and S.-H. Park. (2012). “Preference-based reinforcement learning: a formal framework and a policy iteration algorithm”. *Machine learning*. 89(1-2): 123–156.

Garcıa, J. and F. Fernández. (2015). “A Comprehensive Survey on Safe Reinforcement Learning”. *Journal of Machine Learning Research*. 16(1): 1437–1480.

Ghasemipour, S. K. S., R. Zemel, and S. Gu. (2020). “A divergence minimization perspective on imitation learning methods”. In: *Conference on Robot Learning*. PMLR. 1259–1277.

Gibson, J. J. (1977). “The theory of affordances”. *Hilldale, USA*. 1(2): 67–82.

Goecks, V. G., G. M. Gremillion, V. J. Lawhern, J. Valasek, and N. R. Waytowich. (2019). "Efficiently Combining Human Demonstrations and Interventions for Safe Training of Autonomous Systems in Real-Time". In: *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. AAAI'19/IAAI'19/EAAI'19*. Honolulu, Hawaii, USA: AAAI Press. DOI: 10.1609/aaai.v33i01.33012462.

Goodfellow, I., Y. Bengio, and A. Courville. (2016). *Deep learning*. MIT press.

Griffith, S., K. Subramanian, J. Scholz, C. L. Isbell, and A. Thomaz. (2013). "Policy Shaping: Integrating Human Feedback with Reinforcement Learning". In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. NIPS'13*. Lake Tahoe, Nevada: Curran Associates Inc. 2625–2633.

Grizou, J., M. Lopes, and P.-Y. Oudeyer. (2013). "Robot learning simultaneously a task and how to interpret human instructions". In: *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. 1–8. DOI: 10.1109/DevLrn.2013.6652523.

Haario, H., E. Saksman, and J. Tamminen. (2001). "An adaptive Metropolis algorithm". *Bernoulli*: 223–242.

Halford, G. S., R. Baker, J. E. McCredden, and J. D. Bain. (2005). "How many variables can humans process?" *Psychological science*. 16(1): 70–76.

Hammersley, J. and D. Handscomb. (1964). "Monte carlo methods, methuen & co". *Ltd., London*. 40.

Haykin, S. S. (2001). *Neural networks: a comprehensive foundation*. Tsinghua University Press.

He, X., H. Chen, and B. An. (2020). "Learning Behaviors with Uncertain Human Feedback". In: *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*. Vol. 124. *Proceedings of Machine Learning Research*. PMLR. 131–140. URL: https://proceedings.mlr.press/v124/he20a.html.

Hedlund, E., M. Johnson, and M. Gombolay. (2021). "The Effects of a Robot's Performance on Human Teachers for Learning from Demonstration Tasks". In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction. HRI '21*. Boulder, CO, USA: Association for Computing Machinery. 207–215. DOI: 10.1145/3434073.3444664.

Hester, T., M. Vecerik, O. Pietquin, M. Lanctot, T. Schaul, B. Piot, D. Horgan, J. Quan, A. Sendonaris, I. Osband, *et al.* (2018). "Deep q-learning from demonstrations". In: *Thirty-second AAAI conference on artificial intelligence*.

Hindle, B. R., J. W. Keogh, and A. V. Lorimer. (2021). "Inertial-based human motion capture: A technical summary of current processing methodologies for spatiotemporal and kinematic measures". *Applied Bionics and Biomechanics*. 2021.

Ho, M. K., M. L. Littman, F. Cushman, and J. L. Austerweil. (2015). "Teaching with rewards and punishments: Reinforcement or communication?" In: *CogSci*.

Holzinger, A. (2016). "Interactive machine learning for health informatics: when do we need the human-in-the-loop?" *Brain Informatics*. 3(2): 119–131.

Hoque, R., A. Balakrishna, E. Novoseller, A. Wilcox, D. S. Brown, and K. Goldberg. (2022). "ThriftyDAgger: Budget-Aware Novelty and Risk Gating for Interactive Imitation Learning". In: *Proceedings of the 5th Conference on Robot Learning*. Vol. 164. *Proceedings of Machine Learning Research*. PMLR. 598–608. URL: https://proceedings.mlr.press/v164/hoque22a.html.

Hoque, R., A. Balakrishna, C. Putterman, M. Luo, D. S. Brown, D. Seita, B. Thananjeyan, E. Novoseller, and K. Goldberg. (2021). "LazyDAgger: Reducing Context Switching in Interactive Imitation Learning". In: *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*. 502–509. DOI: 10.1109/CASE49439.2021.9551469.

Howard, R. A. (1960). "Dynamic programming and markov processes".

Huang, Y., L. Rozo, J. Silvério, and D. G. Caldwell. (2019). "Kernelized movement primitives". *The International Journal of Robotics Research*. 38(7): 833–852.

Hurtado, J. V., L. Londoño, and A. Valada. (2021). "From Learning to Relearning: A Framework for Diminishing Bias in Social Robot Navigation". *Frontiers in Robotics and AI*. 8. DOI: 10.3389/frobt.2021.650325.

Hussein, A., M. M. Gaber, E. Elyan, and C. Jayne. (2017). "Imitation learning: A survey of learning methods". *ACM Computing Surveys (CSUR)*. 50(2): 1–35.

Ibarz, B., J. Leike, T. Pohlen, G. Irving, S. Legg, and D. Amodei. (2018). "Reward Learning from Human Preferences and Demonstrations in Atari". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems. NIPS'18*. Montréal, Canada: Curran Associates Inc. 8022–8034.

Isbell, C. and C. Shelton. (2001). "Cobot: A Social Reinforcement Learning Agent". In: *Advances in Neural Information Processing Systems*. Vol. 14. MIT Press. URL: https://proceedings.neurips.cc/paper/2001/hash/92bbd31f8e0e43a7da8a6295b251725f-Abstract.html (accessed on 04/05/2022).

Jain, A., S. Sharma, T. Joachims, and A. Saxena. (2015). "Learning preferences for manipulation tasks from online coactive feedback". *The International Journal of Robotics Research*. 34(10): 1296–1313.

Jain, A., B. Wojcik, T. Joachims, and A. Saxena. (2013). "Learning Trajectory Preferences for Manipulators via Iterative Improvement". In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1. NIPS'13*. Lake Tahoe, Nevada: Curran Associates Inc. 575–583.

James, S., Z. Ma, D. R. Arrojo, and A. J. Davison. (2020). "RLBench: The Robot Learning Benchmark amp; Learning Environment". *IEEE Robotics and Automation Letters*. 5(2): 3019–3026. DOI: 10.1109/LRA.2020.2974707.

Jang, E., A. Irpan, M. Khansari, D. Kappler, F. Ebert, C. Lynch, S. Levine, and C. Finn. (2022). "BC-Z: Zero-Shot Task Generalization with Robotic Imitation Learning". In: *Proceedings of the 5th Conference on Robot Learning*. Vol. 164. *Proceedings of Machine Learning Research*. PMLR. 991–1002. URL: https://proceedings.mlr.press/v164/jang22a.html.

Jauhri, S., C. Celemin, and J. Kober. (2021). "Interactive Imitation Learning in State-Space". In: *Proceedings of the 2020 Conference on Robot Learning*. Vol. 155. *Proceedings of Machine Learning Research*. PMLR. 682–692. URL: https://proceedings.mlr.press/v155/jauhri21a.html.

Jiang, L., S. Liu, and C. Chen. (2019). "Recent research advances on interactive machine learning". *Journal of Visualization*. 22(2): 401–417.

Kaelbling, L. P., M. L. Littman, and A. R. Cassandra. (1998). "Planning and acting in partially observable stochastic domains". *Artificial Intelligence*. 101(1): 99–134. DOI: 10.1016/S0004-3702(98)00023-X.

Kahn, G., P. Abbeel, and S. Levine. (2021). "LaND: Learning to Navigate From Disengagements". *IEEE Robotics and Automation Letters*. 6(2): 1872–1879. DOI: 10.1109/LRA.2021.3060404.

Kamohara, S., H. Takagi, and T. Takeda. (1997). "Control rule acquisition for an arm wrestling robot". In: *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*. Vol. 5. IEEE. 4227–4231.

Kaplan, F., P.-Y. Oudeyer, E. Kubinyi, and A. Miklósi. (2002). "Robotic clicker training". *Robotics and Autonomous Systems*. 38(3): 197–206. DOI: https://doi.org/10.1016/S0921-8890(02)00168-9.

Ke, L., S. Choudhury, M. Barnes, W. Sun, G. Lee, and S. Srinivasa. (2020). "Imitation learning as f-divergence minimization". In: *International Workshop on the Algorithmic Foundations of Robotics*. Springer. 313–329.

Kelly, M., C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer. (2019). "HG-DAgger: Interactive Imitation Learning with Human Experts". In: *2019 International Conference on Robotics and Automation (ICRA)*. 8077–8083. DOI: 10.1109/ICRA.2019.8793698.

Khetarpal, K., Z. Ahmed, G. Comanici, D. Abel, and D. Precup. (2020). "What can I do here? A Theory of Affordances in Reinforcement Learning". In: *International Conference on Machine Learning*. PMLR. 5243–5253.

Kimble, K., K. Van Wyk, J. Falco, E. Messina, Y. Sun, M. Shibata, W. Uemura, and Y. Yokokohji. (2020). "Benchmarking Protocols for Evaluating Small Parts Robotic Assembly Systems". *IEEE Robotics and Automation Letters*. 5(2): 883–889. DOI: 10.1109/LRA.2020.2965869.

Knox, W. B., C. Breazeal, and P. Stone. (2012). "Learning from feedback on actions past and intended". In: *In Proceedings of 7th ACM/IEEE International Conference on Human-Robot Interaction, Late-Breaking Reports Session (HRI 2012)*.

Knox, W. B. and P. Stone. (2008). "Tamer: Training an agent manually via evaluative reinforcement". In: *Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on*. IEEE. 292–297.

Knox, W. B. and P. Stone. (2010). "Combining manual feedback with subsequent MDP reward signals for reinforcement learning". In: *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems. 5–12.

Knox, W. B. and P. Stone. (2012). "Reinforcement learning from simultaneous human and MDP reward". In: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems. 475–482.

Knox, W. B. and P. Stone. (2009). "Interactively Shaping Agents via Human Reinforcement: The TAMER Framework". In: *Proceedings of the Fifth International Conference on Knowledge Capture. K-CAP '09*. Redondo Beach, California, USA: Association for Computing Machinery. 9–16. DOI: 10.1145/1597735.1597738.

Knox, W. B. and P. Stone. (2013). "Learning Non-Myopically from Human-Generated Reward". In: *Proceedings of the 2013 International Conference on Intelligent User Interfaces. IUI '13*. Santa Monica, California, USA: Association for Computing Machinery. 191–202. DOI: 10.1145/2449396.2449422.

Knox, W. B. and P. Stone. (2015). "Framing reinforcement learning from human reward: Reward positivity, temporal discounting, episodicity, and performance". *Artificial Intelligence*. 225: 24–50. DOI: https://doi.org/10.1016/j.artint.2015.03.009.

Knox, W. B., P. Stone, and C. Breazeal. (2013). "Training a Robot via Human Feedback: A Case Study". In: *Proceedings of the 5th International Conference on Social Robotics - Volume 8239. ICSR 2013*. Bristol, UK: Springer-Verlag. 460–470. DOI: 10.1007/978-3-319-02675-6_46.

Kober, J. and J. Peters. (2008). "Policy Search for Motor Primitives in Robotics". 21. URL: https://proceedings.neurips.cc/paper/2008/file/7647966b7343c29048673252e490f736-Paper.pdf.

Koert, D., M. Kircher, V. Salikutluk, C. D'Eramo, and J. Peters. (2020). "Multi-Channel Interactive Reinforcement Learning for Sequential Tasks". *Frontiers in Robotics and AI*. 7. DOI: 10.3389/frobt.2020.00097.

Koert, D., J. Pajarinen, A. Schotschneider, S. Trick, C. Rothkopf, and J. Peters. (2019). "Learning intention aware online adaptation of movement primitives". *IEEE Robotics and Automation Letters*. 4(4): 3719–3726.

Koppol, P., H. Admoni, and R. Simmons. (2021). "Interaction considerations in learning from humans". In: *International Joint Conference on Artificial Intelligence (IJCAI)*.

Krening, S., B. Harrison, K. M. Feigh, C. L. Isbell, M. Riedl, and A. Thomaz. (2017). "Learning From Explanations Using Sentiment and Advice in RL". *IEEE Transactions on Cognitive and Developmental Systems*. 9(1): 44–55. DOI: 10.1109/TCDS.2016.2628365.

Kulak, T., H. Girgin, J.-M. Odobez, and S. Calinon. (2021). "Active learning of Bayesian probabilistic movement primitives". *IEEE Robotics and Automation Letters*. 6(2): 2163–2170.

Kulesza, T., S. Amershi, R. Caruana, D. Fisher, and D. Charles. (2014). "Structured labeling for facilitating concept evolution in machine learning". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3075–3084.

Kuniavsky, M. (2003). *Observing the user experience: a practitioner's guide to user research*. Elsevier.

Laskey, M., C. Chuck, J. Lee, J. Mahler, S. Krishnan, K. Jamieson, A. Dragan, and K. Goldberg. (2017a). "Comparing human-centric and robot-centric sampling for robot deep learning from demonstrations". In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. 358–365. DOI: 10.1109/ICRA.2017.7989046.

Laskey, M., J. Lee, R. Fox, A. Dragan, and K. Goldberg. (2017b). "Dart: Noise injection for robust imitation learning". In: *Conference on robot learning*. PMLR. 143–156.

Laskey, M., S. Staszak, W. Y.-S. Hsieh, J. Mahler, F. T. Pokorny, A. D. Dragan, and K. Goldberg. (2016). "SHIV: Reducing supervisor burden in dagger using support vectors for efficient learning from demonstrations in high dimensional state spaces". In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 462–469.

Le, H., N. Jiang, A. Agarwal, M. Dudik, Y. Yue, and H. Daumé III. (2018). "Hierarchical imitation and reinforcement learning". In: *International conference on machine learning*. PMLR. 2917–2926.

Lee, J. (2017). "A survey of robot learning from demonstrations for human-robot collaboration". *arXiv preprint arXiv:1710.08789*.

Lee, Y., E. S. Hu, Z. Yang, and J. J. Lim. (2020). "To Follow or not to Follow: Selective Imitation Learning from Observations". In: *Proceedings of the Conference on Robot Learning*. Vol. 100. *Proceedings of Machine Learning Research*. PMLR. 11–23. URL: https://proceedings.mlr.press/v100/lee20a.html.

León, A., E. F. Morales, L. Altamirano, and J. R. Ruiz. (2011). "Teaching a Robot to Perform Task through Imitation and On-Line Feedback". In: *Proceedings of the 16th Iberoamerican Congress Conference on Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP'11*. Pucón, Chile: Springer-Verlag. 549–556. DOI: 10.1007/978-3-642-25085-9_65.

Levine, S., A. Kumar, G. Tucker, and J. Fu. (2020). "Offline reinforcement learning: Tutorial, review, and perspectives on open problems". *arXiv preprint arXiv:2005.01643*.

Lewis, M. A., A. H. Fagg, A. Solidum, *et al.* (1992). "Genetic programming approach to the construction of a neural network for control of a walking robot." In: *ICRA*. Citeseer. 2618–2623.

Li, G., R. Gomez, K. Nakamura, and B. He. (2019a). "Human-centered reinforcement learning: A survey". *IEEE Transactions on Human-Machine Systems*. 49(4): 337–349.

Li, G., S. Whiteson, W. B. Knox, and H. Hung. (2016). "Using informative behavior to increase engagement while learning from human reward". *Autonomous agents and multi-agent systems*. 30(5): 826–848.

Li, G., M. Mueller, V. M. Casser, N. Smith, D. Michels, and B. Ghanem. (2019b). "OIL: Observational Imitation Learning". In: *Proceedings of Robotics: Science and Systems*. FreiburgimBreisgau, Germany. DOI: 10.15607/RSS.2019.XV.005.

Liese, F. and I. Vajda. (2006). "On divergences and informations in statistics and information theory". *IEEE Transactions on Information Theory*. 52(10): 4394–4412.

Lin, J., Z. Ma, R. Gomez, K. Nakamura, B. He, and G. Li. (2020). "A Review on Interactive Reinforcement Learning From Human Social Feedback". *IEEE Access*. 8: 120757–120765.

Lin, L.-J. (1992). "Self-Improving Reactive Agents Based on Reinforcement Learning, Planning and Teaching". *Machine Learning*. 8(3): 293–321. DOI: 10.1007/BF00992699.

Lin, Y., A. S. Wang, E. Undersander, and A. Rai. (2022). "Efficient and Interpretable Robot Manipulation With Graph Neural Networks". *IEEE Robotics and Automation Letters*. 7(2): 2740–2747. DOI: 10.1109/LRA.2022.3143518.

Loftin, R., J. MacGlashan, B. Peng, M. Taylor, M. Littman, J. Huang, and D. Roberts. (2014). "A Strategy-Aware Technique for Learning Behaviors from Discrete Human Feedback". In: vol. 28. No. 1. DOI: 10.1609/aaai.v28i1.8839.

Loftin, R., B. Peng, J. MacGlashan, M. L. Littman, M. E. Taylor, J. Huang, and D. L. Roberts. (2016). "Learning behaviors via human-delivered discrete feedback: modeling implicit feedback strategies to speed up learning". *Autonomous agents and multi-agent systems*. 30(1): 30–59.

Londoño, L., A. Röfer, T. Welschehold, and A. Valada. (2022). "Doing Right by Not Doing Wrong in Human-Robot Collaboration". arXiv: 2202.02654 [cs.RO].

Losey, D. P., A. Bajcsy, M. K. O'Malley, and A. D. Dragan. (2022). "Physical interaction as communication: Learning robot objectives online from human corrections". *The International Journal of Robotics Research*. 41(1): 20–44. DOI: 10.1177/02783649211050958.

Lund, H., O. Miglino, L. Pagliarini, A. Billard, and A. Ijspeert. (1998). "Evolutionary robotics-a children's game". In: *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*. 154–158. DOI: 10.1109/ICEC.1998.699493.

Luo, J., O. Sushkov, R. Pevceviciute, W. Lian, C. Su, M. Vecerik, N. Ye, S. Schaal, and J. Scholz. (2021). "Robust Multi-Modal Policies for Industrial Assembly via Reinforcement Learning and Demonstrations: A Large-Scale Study". In: *Robotics: Science and Systems XVII, 2021*.

MacGlashan, J., M. K. Ho, R. Loftin, B. Peng, G. Wang, D. L. Roberts, M. E. Taylor, and M. L. Littman. (2017). "Interactive Learning from Policy-Dependent Human Feedback". In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 2285–2294. URL: https://proceedings.mlr.press/v70/macglashan17a.html.

MacGlashan, J., M. Littman, R. Loftin, B. Peng, D. Roberts, and M. E. Taylor. (2014). "Training an agent to ground commands with reward and punishment". In: *Proceedings of the AAAI Machine Learning for Interactive Systems Workshop*. 6–12.

Maclin, R. and J. W. Shavlik. (1994). *Incorporating Advice into Agents That Learn from Reinforcements*. University of Wisconsin-Madison. Computer Sciences Department.

Maeda, G., M. Ewerton, T. Osa, B. Busch, and J. Peters. (2017). "Active Incremental Learning of Robot Movement Primitives". In: *Proceedings of the 1st Annual Conference on Robot Learning*. Vol. 78. *Proceedings of Machine Learning Research*. PMLR. 37–46. URL: https://proceedings.mlr.press/v78/maeda17a.html.

Mahmood, A. (2017). "Incremental off-policy reinforcement learning algorithms".

Mandlekar, A., J. Booher, M. Spero, A. Tung, A. Gupta, Y. Zhu, A. Garg, S. Savarese, and L. Fei-Fei. (2019). "Scaling Robot Supervision to Hundreds of Hours with RoboTurk: Robotic Manipulation Dataset through Human Reasoning and Dexterity". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 1048–1055. DOI: 10.1109/IROS40897.2019.8968114.

Mandlekar, A., D. Xu, R. Martín-Martín, Y. Zhu, L. Fei-Fei, and S. Savarese. (2020). "Human-in-the-Loop Imitation Learning using Remote Teleoperation". DOI: 10.48550/ARXIV.2012.06733.

Mandlekar, A., Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay, S. Savarese, and L. Fei-Fei. (2018). "ROBOTURK: A Crowdsourcing Platform for Robotic Skill Learning through Imitation". In: *Proceedings of The 2nd Conference on Robot Learning*. Vol. 87. *Proceedings of Machine Learning Research*. PMLR. 879–893. URL: https://proceedings.mlr.press/v87/mandlekar18a.html.

Marta, D., C. Pek, G. I. Melsión, J. Tumova, and I. Leite. (2021). "Human-Feedback Shield Synthesis for Perceived Safety in Deep Reinforcement Learning". *IEEE Robotics and Automation Letters*. 7(1): 406–413.

McCarthy, J. (1958). "Programs with Common Sense". In: RLE and MIT computation center Cambridge, MA, USA.

Menda, K., K. Driggs-Campbell, and M. J. Kochenderfer. (2017). "DropoutDAgger: A Bayesian Approach to Safe Imitation Learning". DOI: 10.48550/ARXIV.1709.06166.

Menda, K., K. Driggs-Campbell, and M. J. Kochenderfer. (2019). "EnsembleDAgger: A Bayesian Approach to Safe Imitation Learning". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 5041–5048. DOI: 10.1109/IROS40897.2019.8968287.

Mericli, C. (2011). "Multi-Resolution Model Plus Correction Paradigm for Task and Skill Refinement on Autonomous Robots". *PhD thesis*. Citeseer.

Meriçli, Ç. and M. Veloso. (2011). "Improving Biped Walk Stability Using Real-Time Corrective Human Feedback". In: *RoboCup 2010: Robot Soccer World Cup XIV*. Berlin, Heidelberg: Springer Berlin Heidelberg. 194–205.

Meriçli, Ç., M. Veloso, and H. L. Akin. (2010). "Complementary humanoid behavior shaping using corrective demonstration". In: *2010 10th IEEE-RAS International Conference on Humanoid Robots*. 334–339. DOI: 10.1109/ICHR.2010.5686326.

Meriçli, Ç., M. Veloso, and H. L. Akın. (2011). "Task Refinement for Autonomous Robots Using Complementary Corrective Human Feedback". *International Journal of Advanced Robotic Systems*. 8(2): 16. DOI: 10.5772/10575.

Mészáros, A., G. Franzese, and J. Kober. (2022). "Learning to Pick at Non-Zero-Velocity From Interactive Demonstrations". *IEEE Robotics and Automation Letters*. 7(3): 6052–6059. DOI: 10.1109/LRA.2022.3165531.

Mitsunaga, N., C. Smith, T. Kanda, H. Ishiguro, and N. Hagita. (2008). "Adapting robot behavior for human–robot interaction". *IEEE Transactions on Robotics*. 24(4): 911–916.

Mohseni, S., N. Zarei, and E. Ragan. (2019). "A Multidisciplinary survey and framework for design and evaluation of explainable AI systems. arXiv". *Human-Computer Interaction*.

Müller, M., V. Casser, J. Lahoud, N. Smith, and B. Ghanem. (2018). "Sim4cv: A photo-realistic simulator for computer vision applications". *International Journal of Computer Vision*. 126(9): 902–919.

Myers, V., E. Biyik, N. Anari, and D. Sadigh. (2022). "Learning Multimodal Rewards from Rankings". In: *Proceedings of the 5th Conference on Robot Learning*. Vol. 164. *Proceedings of Machine Learning Research*. PMLR. 342–352. URL: https://proceedings.mlr.press/v164/myers22a.html.

Najar, A. and M. Chetouani. (2021). "Reinforcement Learning with Human Advice: A Survey". *Frontiers in Robotics and AI*. 8.

Najar, A., O. Sigaud, and M. Chetouani. (2016). "Training a robot with evaluative feedback and unlabeled guidance signals". In: *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*. IEEE. 261–266.

Najar, A., O. Sigaud, and M. Chetouani. (2020). "Interactively shaping robot behaviour with unlabeled human instructions". *Autonomous Agents and Multi-Agent Systems*. 34(2): 1–35.

Nehaniv, C. L., K. Dautenhahn, *et al.* (2002). "The correspondence problem". *Imitation in animals and artifacts*. 41.

Ng, A. Y., D. Harada, and S. Russell. (1999). "Policy Invariance under Reward Transformations: Theory and Application to Reward Shaping". In: *Icml*. Vol. 99. 278–287.

Ng, A. Y. and S. J. Russell. (2000). "Algorithms for inverse reinforcement learning." In: *Icml*. 663–670.

Ngo, H., M. Luciw, J. Nagi, A. Forster, J. Schmidhuber, and N. A. Vien. (2014). "Efficient interactive multiclass learning from binary feedback". *ACM Transactions on Interactive Intelligent Systems (TiiS)*. 4(3): 12.

Nicolescu, M. N. and M. J. Mataric. (2003). "Natural Methods for Robot Task Learning: Instructive Demonstrations, Generalization and Practice". In: *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems. AAMAS '03*. Melbourne, Australia: Association for Computing Machinery. 241–248. DOI: 10.1145/860575.860614.

Nojima, Y., F. Kojima, and N. Kubota. (2003). "Trajectory generation for human-friendly behavior of partner robot using fuzzy evaluating interactive genetic algorithm". In: *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Computational Intelligence in Robotics and Automation for the New Millennium (Cat. No. 03EX694)*. Vol. 1. IEEE. 306–311.

Osa, T., J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, and J. Peters. (2018). "An algorithmic perspective on imitation learning". *arXiv preprint arXiv:1811.06711*.

Palan, M., G. Shevchuk, N. Charles Landolfi, and D. Sadigh. (2019). "Learning Reward Functions by Integrating Human Demonstrations and Preferences". In: *Robotics: Science and Systems*.

Paraschos, A., C. Daniel, J. R. Peters, and G. Neumann. (2013). "Probabilistic movement primitives". In: *Advances in neural information processing systems*. 2616–2624.

Paull, L., J. Tani, H. Ahn, J. Alonso-Mora, L. Carlone, M. Cap, Y. F. Chen, C. Choi, J. Dusek, Y. Fang, *et al.* (2017). "Duckietown: an open, inexpensive and flexible platform for autonomy education and research". In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE. 1497–1504.

Peng, B., J. MacGlashan, R. Loftin, M. L. Littman, D. L. Roberts, and M. E. Taylor. (2016). "A need for speed: Adapting agent action speed to improve task learning from non-expert humans". In: *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems*.

Pérez-Dattari, R., B. Brito, O. de Groot, J. Kober, and J. Alonso-Mora. (2022). "Visually-guided motion planning for autonomous driving from interactive demonstrations". *Engineering Applications of Artificial Intelligence*. 116. DOI: https://doi.org/10.1016/j.engappai.2022.105277.

Pérez-Dattari, R., C. Celemin, G. Franzese, J. Ruiz-del-Solar, and J. Kober. (2020). "Interactive learning of temporal features for control: Shaping policies and state representations from human feedback". *IEEE Robotics & Automation Magazine*. 27(2): 46–54.

Pérez-Dattari, R., C. Celemin, J. Ruiz-del-Solar, and J. Kober. (2018). "Interactive learning with corrective feedback for policies based on deep neural networks". In: *International Symposium on Experimental Robotics*. Springer. 353–363.

Pérez-Dattari, R., C. Celemin, J. Ruiz-del-Solar, and J. Kober. (2019). "Continuous control for high-dimensional state spaces: An interactive learning approach". In: *2019 International Conference on Robotics and Automation (ICRA)*. IEEE. 7611–7617.

Pilarski, P. M., M. R. Dawson, T. Degris, F. Fahimi, J. P. Carey, and R. S. Sutton. (2011). "Online human training of a myoelectric prosthesis controller via actor-critic reinforcement learning". In: *2011 IEEE International Conference on Rehabilitation Robotics*. 1–7. DOI: 10.1109/ICORR.2011.5975338.

Prakash, A., A. Behl, E. Ohn-Bar, K. Chitta, and A. Geiger. (2020). "Exploring data aggregation in policy learning for vision-based urban autonomous driving". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11763–11773. DOI: 10.1109/CVPR42600.2020.01178.

Precup, D. (2000). *Temporal abstraction in reinforcement learning*. University of Massachusetts Amherst.

Pryor, K. (1999). "Clicker training for dogs. Waltham, MA".

Rasmussen, C. E. and C. K. I. Williams. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.

Ravichandar, H., A. S. Polydoros, S. Chernova, and A. Billard. (2020). "Recent advances in robot learning from demonstration". *Annual Review of Control, Robotics, and Autonomous Systems*. 3: 297–330.

Reddy, S., A. D. Dragan, and S. Levine. (2019). "SQIL: Imitation Learning via Regularized Behavioral Cloning". *CoRR*. abs/1905.11108. URL: http://arxiv.org/abs/1905.11108.

Rohmer, E., S. P. Singh, and M. Freese. (2013). "V-REP: A versatile and scalable robot simulation framework". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 1321–1326.

Ross, S. and J. A. Bagnell. (2014). "Reinforcement and imitation learning via interactive no-regret learning". *arXiv preprint arXiv:1406.5979*.

Ross, S. and D. Bagnell. (2010). "Efficient reductions for imitation learning". In: *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings. 661–668.

Ross, S., G. Gordon, and D. 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*. 627–635.

Rubinstein, R. Y. and D. P. Kroese. (2016). *Simulation and the Monte Carlo method*. Vol. 10. John Wiley & Sons.

Rummery, G. A. and M. Niranjan. (1994). *On-line Q-learning using connectionist systems*. Vol. 37. Citeseer.

Russell, S. J. and P. Norvig. (2016). *Artificial Intelligence : A Modern Approach*. Malaysia and Pearson Education Limited.

Sadigh, D., A. Dragan, S. Sastry, and S. Seshia. (2017). "Active Preference-Based Learning of Reward Functions". In: *Proceedings of Robotics: Science and Systems*. Cambridge, Massachusetts. DOI: 10.15607/RSS.2017.XIII.053.

Samuel, A. L. (1959). "Some Studies in Machine Learning Using the Game of Checkers". *IBM Journal of Research and Development*. 3(3): 210–229. DOI: 10.1147/rd.33.0210.

Samuel, A. L. (1967). "Some Studies in Machine Learning Using the Game of Checkers. II–Recent Progress". *IBM Journal of Research and Development*. 11(6): 601–617. DOI: 10.1147/rd.116.0601.

Saran, A., R. Zhang, E. S. Short, and S. Niekum. (2021). "Efficiently Guiding Imitation Learning Agents with Human Gaze". In: *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. 1109–1117.

Saveriano, M., F. J. Abu-Dakka, A. Kramberger, and L. Peternel. (2021). "Dynamic Movement Primitives in Robotics: A Tutorial Survey". DOI: 10.48550/ARXIV.2102.03861.

Savva, M., A. Kadian, O. Maksymets, Y. Zhao, E. Wijmans, B. Jain, J. Straub, J. Liu, V. Koltun, J. Malik, D. Parikh, and D. Batra. (2019). "Habitat: A Platform for Embodied AI Research". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Scholten, J., D. Wout, C. Celemin, and J. Kober. (2019). "Deep Reinforcement Learning with Feedback-based Exploration". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. 803–808.

Schroecker, Y., H. Ben Amor, and A. Thomaz. (2016). "Directing Policy Search with Interactively Taught Via-Points". In: *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems. AAMAS '16*. Singapore, Singapore: International Foundation for Autonomous Agents and Multiagent Systems. 1052–1059.

Schulman, J., Y. Duan, J. Ho, A. Lee, I. Awwal, H. Bradlow, J. Pan, S. Patil, K. Goldberg, and P. Abbeel. (2014). "Motion planning with sequential convex optimization and convex collision checking". *The International Journal of Robotics Research*. 33(9): 1251–1270.

Settles, B. (2009). "Active learning literature survey".

Shah, S., D. Dey, C. Lovett, and A. Kapoor. (2018). "Airsim: High-fidelity visual and physical simulation for autonomous vehicles". In: *Field and service robotics*. Springer. 621–635.

Shridhar, M., D. Mittal, and D. Hsu. (2020). "INGRESS: Interactive visual grounding of referring expressions". *The International Journal of Robotics Research*. 39(2-3): 217–232. DOI: 10.1177/0278364919897133.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. (2016). "Mastering the Game of Go with Deep Neural Networks and Tree Search". *Nature*. 529(7587): 484–489. DOI: 10.1038/nature16961.

Sinha, S., A. Mandlekar, and A. Garg. (2022). "S4RL: Surprisingly Simple Self-Supervision for Offline Reinforcement Learning in Robotics". In: *Proceedings of the 5th Conference on Robot Learning*. Vol. 164. *Proceedings of Machine Learning Research*. PMLR. 907–917. URL: https://proceedings.mlr.press/v164/sinha22a.html.

Smith, J. R. (1991). "Designing biomorphs with an interactive genetic algorithm." In: *ICGA*. Citeseer. 535–538.

Spencer, J., S. Choudhury, M. Barnes, M. Schmittle, M. Chiang, P. Ramadge, and S. Srinivasa. (2020). "Learning from interventions". In: *Robotics: Science and Systems (RSS)*.

Sperrle, F., M. El-Assady, G. Guo, R. Borgo, D. H. Chau, A. Endert, and D. Keim. (2021). "A Survey of Human-Centered Evaluations in Human-Centered Machine Learning". In: *Computer Graphics Forum*. Vol. 40. No. 3. Wiley Online Library. 543–568.

Sridharan, M. (2011). "Augmented reinforcement learning for interaction with non-expert humans in agent domains". In: *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*. Vol. 1. IEEE. 424–429.

Stapelberg, B. and K. M. Malan. (2020). "A survey of benchmarks for reinforcement learning algorithms". *South African Computer Journal*. 32(2).

Stulp, F. and O. Sigaud. (2015). "Many regression algorithms, one unified model: A review". *Neural Networks*. 69: 60–79.

Suay, H. B. and S. Chernova. (2011). "Effect of human guidance and state space size on interactive reinforcement learning". In: *RO-MAN, 2011 IEEE*. IEEE. 1–6.

Subramanian, K., C. L. Isbell Jr, and A. L. Thomaz. (2016). "Exploration from demonstration for interactive reinforcement learning". In: *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. 447–456.

Sugiyama, M. (2015). *Introduction to statistical machine learning*. Morgan Kaufmann.

Sun, W., A. Venkatraman, G. J. Gordon, B. Boots, and J. A. Bagnell. (2017). "Deeply AggreVaTeD: Differentiable Imitation Learning for Sequential Prediction". In: *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70. *Proceedings of Machine Learning Research*. PMLR. 3309–3318. URL: https://proceedings.mlr.press/v70/sun17d.html.

Sutton, R. S. and A. G. Barto. (2018). *Reinforcement learning: An introduction*. MIT press.

Szot, A., A. Clegg, E. Undersander, E. Wijmans, Y. Zhao, J. M. Turner, N. D. Maestre, M. Mukadam, D. S. Chaplot, O. Maksymets, *et al.* (2021). "Habitat 2.0: Training Home Assistants to Rearrange their Habitat". In: *Thirty-Fifth Conference on Neural Information Processing Systems*.

Takagi, H. (1998). "Interactive evolutionary computation". In: *Proceedings of the International Conference on Soft Computing and Information/Intelligent Systems*. 41–50.

Takagi, H. (2001). "Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation". *Proceedings of the IEEE*. 89(9): 1275–1296.

Tenorio-Gonzalez, A. C., E. F. Morales, and L. Villaseñor-Pineda. (2010). "Dynamic Reward Shaping: Training a Robot by Voice". In: *Advances in Artificial Intelligence – IBERAMIA 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg. 483–492.

Thomaz, A. and C. Breazeal. (2006). "Adding guidance to interactive reinforcement learning". In: *Proceedings of the Twentieth Conference on Artificial Intelligence (AAAI)*.

Thomaz, A. L. and C. Breazeal. (2007a). "Asymmetric Interpretations of Positive and Negative Human Feedback for a Social Learning Agent". In: *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*. 720–725. DOI: 10.1109/ROMAN.2007.4415180.

Thomaz, A. L. and C. Breazeal. (2007b). "Robot learning via socially guided exploration": 82–87. DOI: 10.1109/DEVLRN.2007.4354078.

Thomaz, A. L. and M. Cakmak. (2009). "Learning about objects with human teachers". In: *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 15–22. DOI: 10.1145/1514095.1514101.

Thomaz, A. L., C. Breazeal, *et al.* (2006). "Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance". In: *Aaai*. Vol. 6. Boston, MA. 1000–1005.

Thomaz, A. L., G. Hoffman, and C. Breazeal. (2005). "Real-Time Interactive Reinforcement Learning for Robots". In: *AAAI 2005 Workshop on Human Comprehensible Machine Learning*. 9–13.

Todorov, E., T. Erez, and Y. Tassa. (2012). "MuJoCo: A physics engine for model-based control". In: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 5026–5033. DOI: 10.1109/IROS.2012.6386109.

Torabi, F., G. Warnell, and P. Stone. (2018). "Behavioral cloning from observation". *arXiv preprint arXiv:1805.01954*.

Toris, R., H. B. Suay, and S. Chernova. (2012). "A practical comparison of three robot learning from demonstration algorithms". In: *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 261–262.

Utgoff, P. (1991). "Two Kinds of Training Information for Evaluation Function Learning". *Computer Science Department Faculty Publication Series*: 193.

Van Der Laan, J. D., A. Heino, and D. De Waard. (1997). "A simple procedure for the assessment of acceptance of advanced transport telematics". *Transportation Research Part C: Emerging Technologies*. 5(1): 1–10.

Vecerik, M., T. Hester, J. Scholz, F. Wang, O. Pietquin, B. Piot, N. Heess, T. Rothörl, T. Lampe, and M. Riedmiller. (2017). "Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards". *arXiv preprint arXiv:1707.08817*.

Vien, N. A. and W. Ertel. (2012). "Reinforcement learning combined with human feedback in continuous state and action spaces". In: *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. 1–6. DOI: 10.1109/DevLrn.2012.6400849.

Vien, N. A., W. Ertel, and T. C. Chung. (2013). "Learning via human feedback in continuous state and action spaces". *Applied intelligence*. 39(2): 267–278.

Vollmer, A.-L. and N. J. Hemion. (2018). "A user study on robot skill learning without a cost function: Optimization of dynamic movement primitives via naive user feedback". *Frontiers in Robotics and AI*. 5: 77.

Ware, M., E. Frank, G. Holmes, M. Hall, and I. H. Witten. (2001). "Interactive machine learning: letting users build classifiers". *International Journal of Human-Computer Studies*. 55(3): 281–292.

Warnell, G., N. Waytowich, V. Lawhern, and P. Stone. (2018). "Deep tamer: Interactive agent shaping in high-dimensional state spaces". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1.

Watkins, C. J. and P. Dayan. (1992). "Q-learning". *Machine learning*. 8(3): 279–292.

Watkins, C. J. C. H. (1989). “Learning from delayed rewards”.

Wenskovitch, J. and C. North. (2020). “Interactive Artificial Intelligence: Designing for the" Two Black Boxes" Problem”. *Computer*. 53(8): 29–39.

Whitehead, S. D. (1991). “A Complexity Analysis of Cooperative Mechanisms in Reinforcement Learning.” In: *AAAI*. 607–613.

Wilde, N., E. Biyik, D. Sadigh, and S. L. Smith. (2022). “Learning Reward Functions from Scale Feedback”. In: *Proceedings of the 5th Conference on Robot Learning*. Vol. 164. *Proceedings of Machine Learning Research*. PMLR. 353–362. URL: https://proceedings.mlr.press/v164/wilde22a.html.

Wilde, N., A. Blidaru, S. L. Smith, and D. Kulić. (2020). “Improving user specifications for robot behavior through active preference learning: Framework and evaluation”. *The International Journal of Robotics Research*. 39(6): 651–667. DOI: 10.1177/0278364920910802.

Williams, E. C., N. Gopalan, M. Rhee, and S. Tellex. (2018). “Learning to Parse Natural Language to Grounded Reward Functions with Weak Supervision”. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 4430–4436. DOI: 10.1109/ICRA.2018.8460937.

Wilson, A., A. Fern, and P. Tadepalli. (2012). “A Bayesian Approach for Policy Learning from Trajectory Preference Queries”. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. NIPS’12*. Lake Tahoe, Nevada: Curran Associates Inc. 1133–1141.

Wout, D., J. Scholten, C. Celemin, and J. Kober. (2019). “Learning Gaussian Policies from Corrective Human Feedback”. DOI: 10.48550/ARXIV.1903.05216.

Wrede, S., C. Emmerich, R. Grünberg, A. Nordmann, A. Swadzba, and J. Steil. (2013). “A User Study on Kinesthetic Teaching of Redundant Robots in Task and Configuration Space”. *J. Hum.-Robot Interact.* 2(1): 56–81. DOI: 10.5898/JHRI.2.1.Wrede.

Wu, X., L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He. (2021). “A Survey of Human-in-the-loop for Machine Learning”. *arXiv preprint arXiv:2108.00941*.

Wulfmeier, M., D. Z. Wang, and I. Posner. (2016). "Watch this: Scalable cost-function learning for path planning in urban environments". In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2089–2095.

Xiao, B., Q. Lu, B. Ramasubramanian, A. Clark, L. Bushnell, and R. Poovendran. (2020). "FRESH: Interactive Reward Shaping in High-Dimensional State Spaces Using Human Feedback". In: *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems. AAMAS '20*. Auckland, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems. 1512–1520.

Xin, D., L. Ma, J. Liu, S. Macke, S. Song, and A. Parameswaran. (2018). "Accelerating human-in-the-loop machine learning: Challenges and opportunities". In: *Proceedings of the second workshop on data management for end-to-end machine learning*. 1–4.

Yang, S., W. Zhang, W. Lu, H. Wang, and Y. Li. (2019). "Learning Actions from Human Demonstration Video for Robotic Manipulation". In: *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 1805–1811. DOI: 10.1109/IROS40897.2019.8968278.

Yanik, P. M., J. Manganelli, J. Merino, A. L. Threatt, J. O. Brooks, K. E. Green, and I. D. Walker. (2014). "A Gesture Learning Interface for Simulated Robot Path Shaping With a Human Teacher". *IEEE Transactions on Human-Machine Systems*. 44(1): 41–54. DOI: 10.1109/TSMC.2013.2291714.

Yu, T., D. Quillen, Z. He, R. Julian, K. Hausman, C. Finn, and S. Levine. (2019). "Meta-World: A Benchmark and Evaluation for Multi-Task and Meta Reinforcement Learning". In: *Conference on Robot Learning (CoRL)*. URL: https://arxiv.org/abs/1910.10897.

Zanzotto, F. M. (2019). "Human-in-the-loop artificial intelligence". *Journal of Artificial Intelligence Research*. 64: 243–252.

Zhang, J. and K. Cho. (2016). "Query-Efficient Imitation Learning for End-to-End Autonomous Driving". DOI: 10.48550/ARXIV.1605.06450.

Zhang, Q., J. Lin, Q. Sha, B. He, and G. Li. (2020). "Deep Interactive Reinforcement Learning for Path Following of Autonomous Underwater Vehicle". *IEEE Access*. 8: 24258–24268. DOI: 10.1109/ACCESS.2020.2970433.

Zhang, Q., L. Zha, J. Lin, D. Tu, M. Li, F. Liang, R. Wu, and X. Lu. (2019a). "A Survey on Deep Learning Benchmarks: Do We Still Need New Ones?" In: *Benchmarking, Measuring, and Optimizing*. Cham: Springer International Publishing. 36–49.

Zhang, R., F. Torabi, L. Guan, D. H. Ballard, and P. Stone. (2019b). "Leveraging human guidance for deep reinforcement learning tasks". *arXiv preprint arXiv:1909.09906*.

Zhifei, S. and E. M. Joo. (2012). "A review of inverse reinforcement learning theory and recent advances". In: *Evolutionary Computation (CEC), 2012 IEEE Congress on*. IEEE. 1–8.

Zhu, Y., J. Wong, A. Mandlekar, and R. Martín-Martín. (2020). "robosuite: A Modular Simulation Framework and Benchmark for Robot Learning". arXiv: 2009.12293 [cs.RO].