

Interactive RL for Robotics

برديا طور انداز

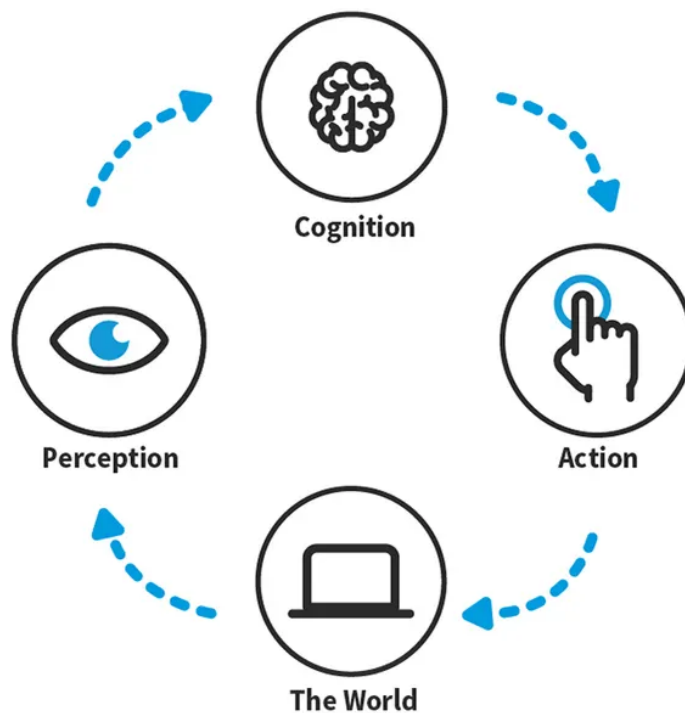
عليرضا عباسي



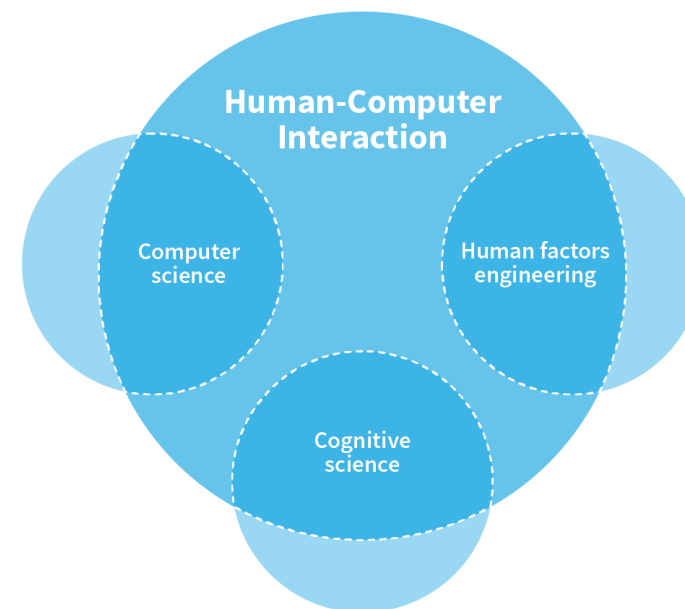


یادگیری تقویتی تعاملی برای رباتیک

- انسان در حلقه (Human-in-the-loop)
- تعامل انسان و کامپیوتر



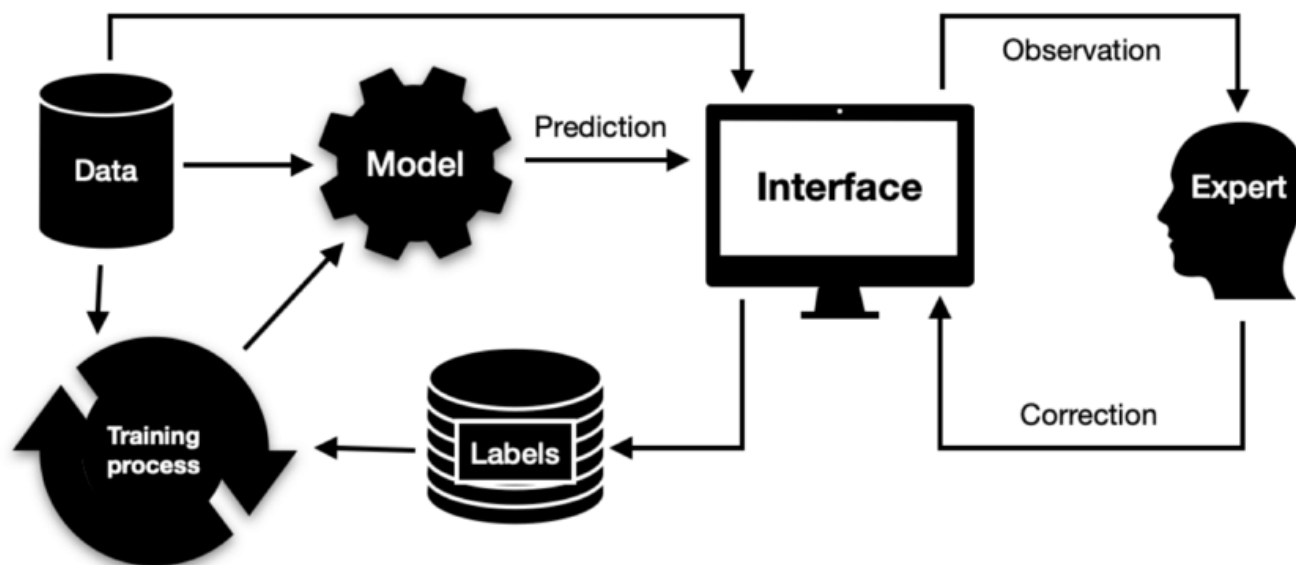
The Multidisciplinary Field of HCI





وظایف مهندسين تعامل انسان و کامپیوتر در زمینه IML

- تعريف تکنیک‌های تعامل انسان و کامپیوتر
- پيدا کردن کاربردهای جديد



ویژگیهای IRL

- سریعتر از Automatic RL
- Generalized نبودن پاسخها
- پیچیدگی تعامل با انسانهای متفاوت

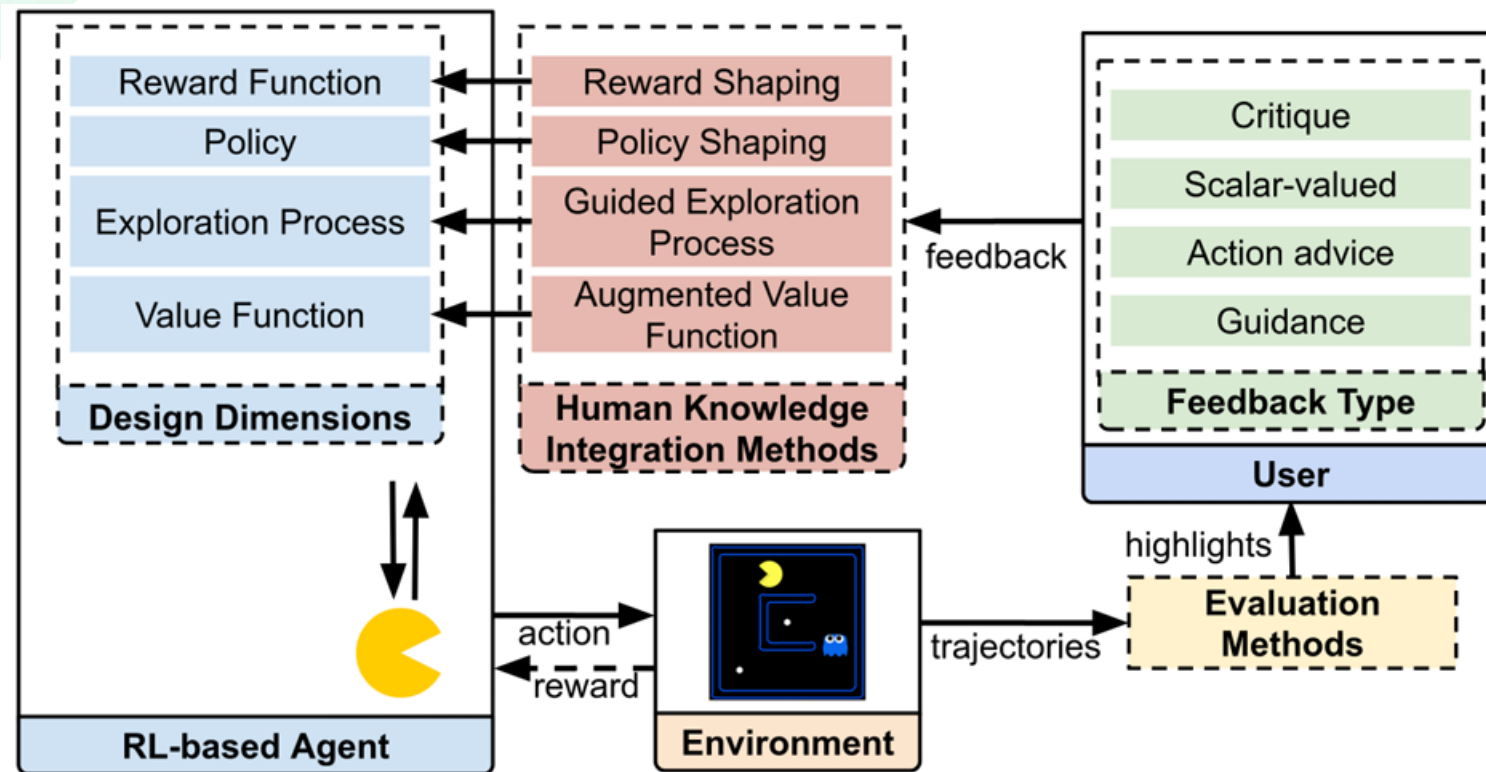
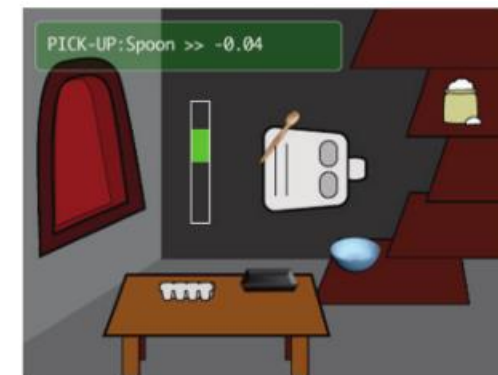
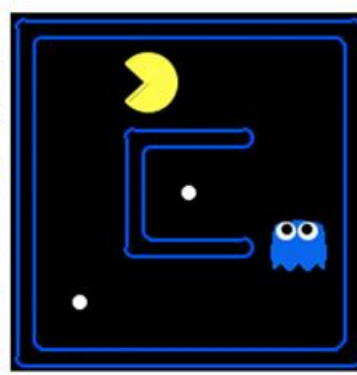
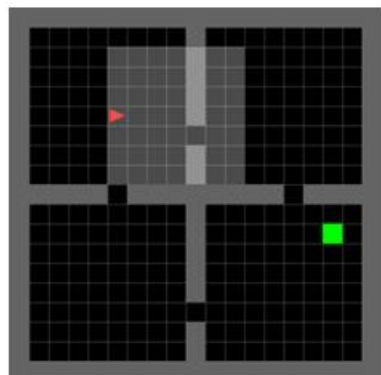


Figure 1. The interactive RL architecture.



محیط‌های آزمایشی یادگیری تقویتی تعاملی

- رباتیک
- هوش مصنوعی در بازی‌ها
- تعامل انسان و کامپیوتر



ابعاد طراحی

$R' = R + F$, where $F : S \times A \times S \rightarrow \mathbb{R}$ is the *shaping reward*

- تابع پاداش: فیدبک انسانی (توسط شخص خبره)، جهت تسریع یادگیری
- سیاست بهینه
- کاوش
- تابع ارزش

Design Dimension	Testbed	Interaction	Initiative	HKI	Feedback	Algorithms
Reward Function	Robot in maze-like environment [10]	FE	Passive	RS using HF + ER	Critique	DQL [67, 77]
	Navigation simulation [12]	GUI	Passive	Advantage Function	Critique	DAC
	Sophie's Kitchen game [105, 103]	GUI	Passive, Active	RS using HF + EF	Critique	QL [109], HRL
	Bowling game [108]	GUI	Passive	RS + HF	Scalar-valued	DQL
	Shopping assistant, GridWorld [76]	GUI	Active	Active IRD	Queries	Model-based RL
	Mario, GridWorld, Soccer simulation [85]	Coding	Passive	PBRs	Heuristic Function	QL, QSL, QL(λ), QSL(λ)
	Navigation simulation [102]	VC	Passive	RS using HF + ER	AcAd	SARSA [86], SARSA(λ)
	Atari, robotics simulation [20]	GUI	Active	RS using HF	Queries	DRL
Policy	GridWorld, TurtleBot robot [71]	GUI, GC	Passive	PS	AcAd	AC(λ) [15, 91]
	GridWorld [56]	VC	Passive	PS	Critique, AcAd	BQL [26]
	Pac-Man, Frogger [37]	GUI	Passive	PS	Critique	BQL
Exploration Process	Pac-Man, Cart-Pole simulation [114]	GUI	Passive	GEP	AcAd	QL
	Simulated cleaning Robot [24, 23]	VC	Passive	GEP	AcAd	SARSA
	Pac-Man [7]	GUI	Active	GEP	AcAd	SARSA(λ)
	Pac-Man [32]	GUI	Active	Myopic Agent	AcAd	QL, QRL
	Sophie's Kitchen game [103]	GUI	Active	ACTG	Guidance	QL
	Street Fighter game [13]	Not apply	Passive	EB using Safe RL	Demonstration	HRL
	Nao Robot [93]	GUI	Passive	ACTG	Guidance	QL
	Nexi robot [54]	AT + CT	Passive	Myopic Agent	AcAd	SARSA(λ)
Value Function	Mountain Car simulation [52]	GUI	Passive	Weighted VF	Demonstration	SARSA(λ)
	Keepaway simulation [101]	GUI	Passive	Weighted VF	Demonstration	SARSA
	Mario, Cart Pole [18]	Not apply	Passive	Initialization of VF	Demonstration	QL(λ)

ابعاد طراحی

- طراحی پاداش:

The reward shaping (RS) method aims to mold the behavior of a learning agent by modifying its reward function to encourage the behavior the RL designer wants.

- طراحی سیاست:

The policy shaping (PS) approach consists of directly molding the policy of a learning agent to fit its behavior to what the RL designer envisions.

- فرآیند کاوش هدایت شده:

Guided exploration process methods aim to minimize the learning procedure by injecting human knowledge to guide the agent's exploration to states with a high reward.

- تابع ارزش تقویت شده:

The procedure to augment a value function consists of combining the value function of the agent with one created from human feedback.

مسائل تحقیقاتی

High-dimensional
Environments

محیط‌های با ابعاد بالا

Lack of Evaluation
Techniques

کمبود روش‌های ارزیابی

Lack of Human-like
Oracles

فقدان مشاوران شبیه به انسان

Modeling Users

مدل‌سازی کاربران

Combining Different
Design Dimensions

ترکیب ابعاد مختلف طراحی

Safe Interactive RL

یادگیری تقویتی تعاملی ایمن

Fast Evaluation of
Behaviors

ارزیابی سریع رفتارها

Explainable Interactive
RL

یادگیری تقویتی تعاملی قابل
توضیح