

# Causality Agents & Foundation Models

Mengyue Yang

PhD Candidate

University College London

Email: [mengyue.yang.20@ucl.ac.uk](mailto:mengyue.yang.20@ucl.ac.uk)



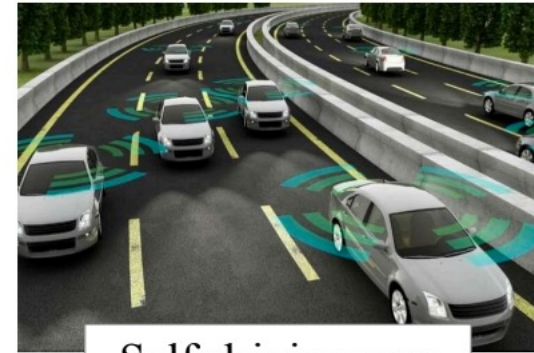
# Agents in decision-making systems



Social media



Healthcare robots



Self-driving cars



Industry production



Smart city

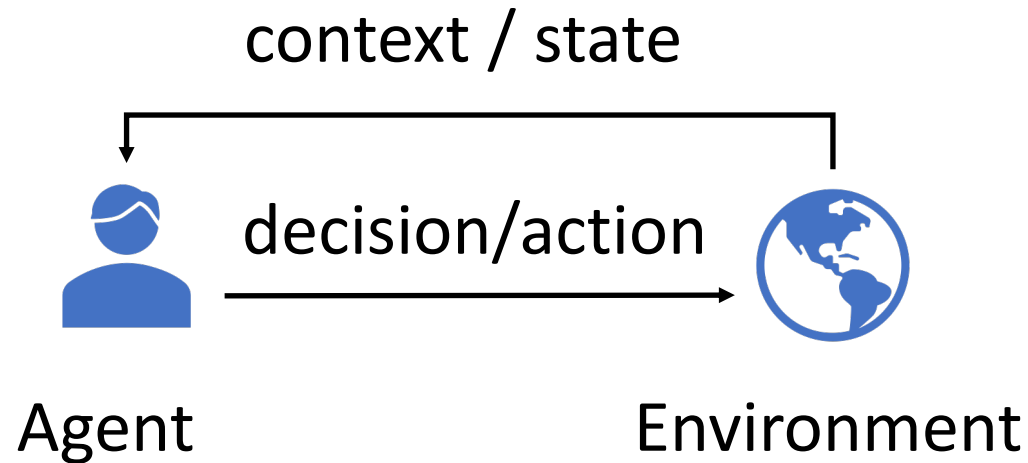


Drone delivery

... some material is borrowed from Chongjie Zhang, Jun Wang, Shimon Whiteson,...

# What is Agents?

The agents have the ability to plan a trajectory of decisions at the beginning or make real-time decisions at each timestep.



It follows Markov Decision Process (MDP) to make the decision.

# Content

- Introduction of Agents.
  - **Agents**
  - LLM Agents
- Causality and Agents
  - Causality inspired decision making
  - Causality and single agent
  - Causality and multi-agent
- Causality and LLM Agents

# Differences: Agents, Prediction, and Generation

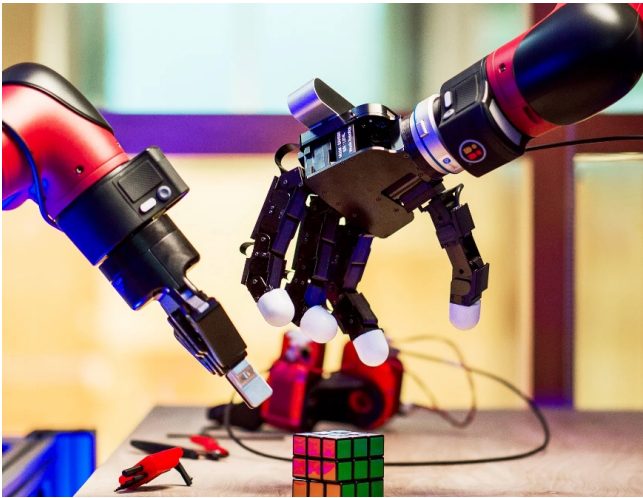
**Agents:** Planning and Making decision.

**Prediction:** Forecasting future events based on current and past data without necessarily taking actions.

**Generation:** Generating data by specific requirements.

# Traditional researches about agent

Robotics, Games, Autonomous driving...



Enabling data-driven machine learning algorithm design such as planning and control.

# Planning and Control for decision making

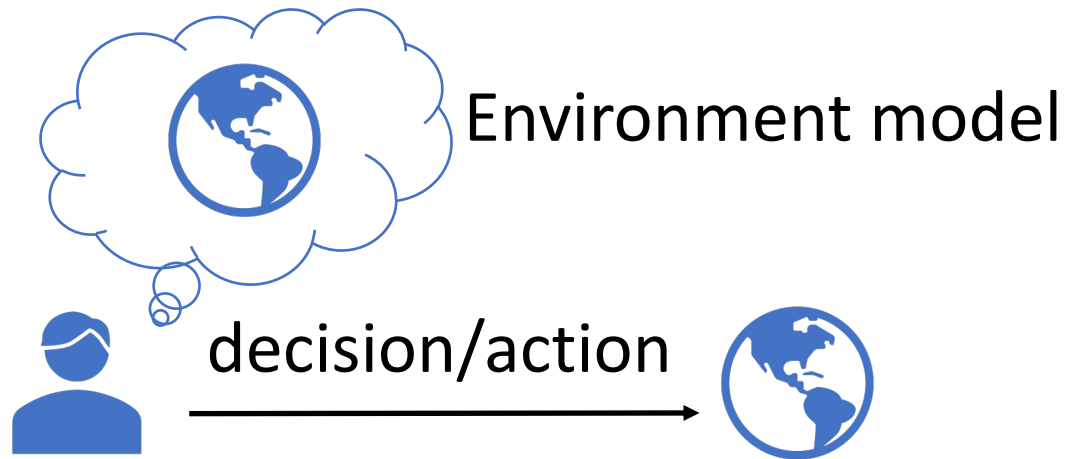
- **Planning:**
  1. Decide the actions for the upcoming timesteps.
  2. These actions, once determined, cannot be changed.
- **Control:**
  1. Only the current action is determined based on past information.
  2. Strategy is adjusted considering feedback.

**Reinforcement Learning (RL) is a popular algorithm to solve the planning and control problem.**

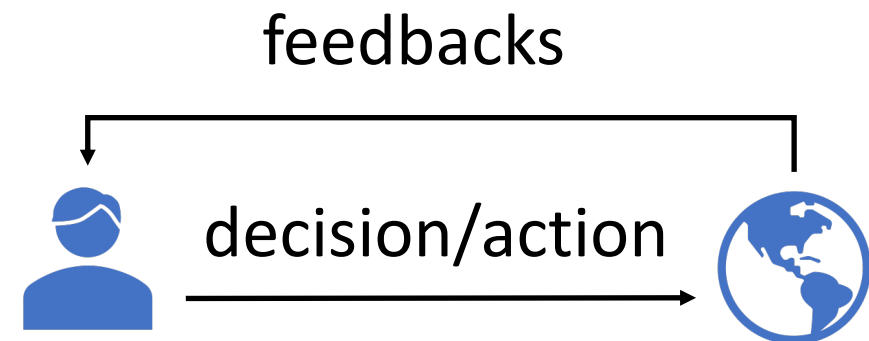
# Algorithms for Planning and Control: RL

## Reinforcement Learning (RL):

- Model-free RL: Making strategy without modelling the system (a type of control algorithm).
- Model-based RL: A type of planning algorithm.



Model-based RL



Model-free RL

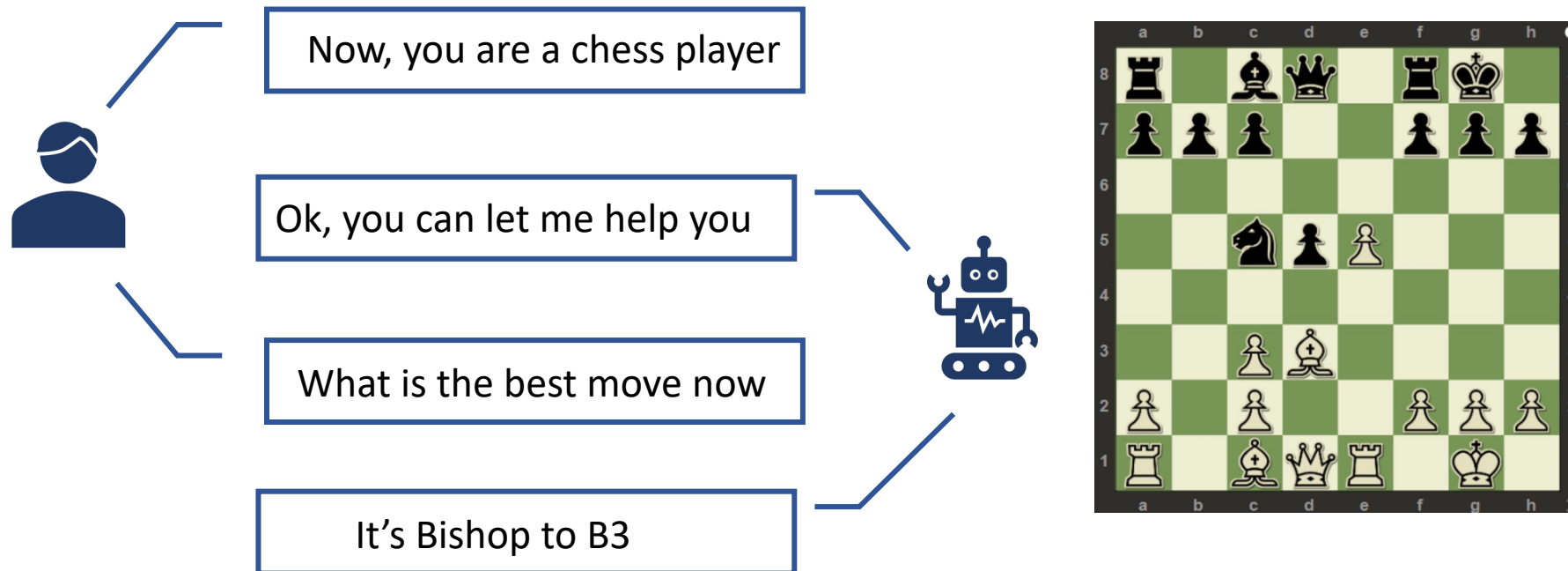


# Content

- Introduction of Agents.
  - Agents
  - **LLM Agents**
- Causality and Agents
  - Causality inspired decision making
  - Causality and single agent
  - Causality and multi-agent
- Causality and LLM Agents

# Foundation Model

- Natural language / multi-modalities generation model.
- A foundational model is a powerful function capable of mimicking a broad class of functions and playing different roles.



# Foundation Model and Agents (LLM Agents)

LLM (Large Language Model) Agent is an intelligent agent where large language models are the core component for reasoning and decision-making.

- **More scenarios:** Any decision scenario expressible in text, such as multi-agent decision making, coding, and mathematical problems.

LLM Agent can perform complex reasoning and decision-making, simulating human thought processes to solve problems.

# Approaches of Foundation Model Agents

- Prompt engineering
- RL tuning
- Expert data collection

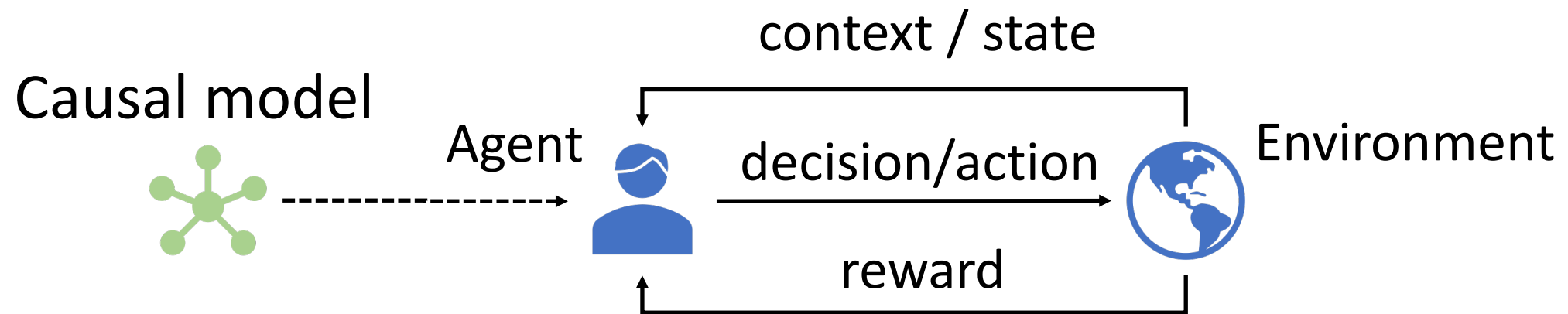
**Challenges:** Many decisions sometimes revert to non-ML methods.

# Content

- Introduction of Agents.
  - Agents
  - LLM Agents
- Causality and Agents
  - **Causality inspired decision making**
  - Causality and single agent
  - Causality and multi-agent
- Causality and LLM Agents

# Causality and Agents: the big picture

- Understanding the factors in the system
- Learning how to take actions



**Understanding**



**Better quality**

# Advantages of causality

- Clarifying Causal information: Identify the key factors and avoid being misled by spurious correlations.
- Enhancing Decision Accuracy and Effectiveness: predict the outcomes make the wisest choices.
- Reducing Decision Risks: Identify potential bad effects and avoid risks of generalization.

# Content

- Introduction of Agents.
  - Agents
  - LLM Agents
- Causality and Agents
  - Causality inspired decision making
  - **Causality and single agent**
  - Causality and multi-agent
- Causality and LLM Agents



# Tasks for causal single agent

- Understanding the world/environment.
- How to make the decision based on causal understanding.

**Understanding** → **Better quality**

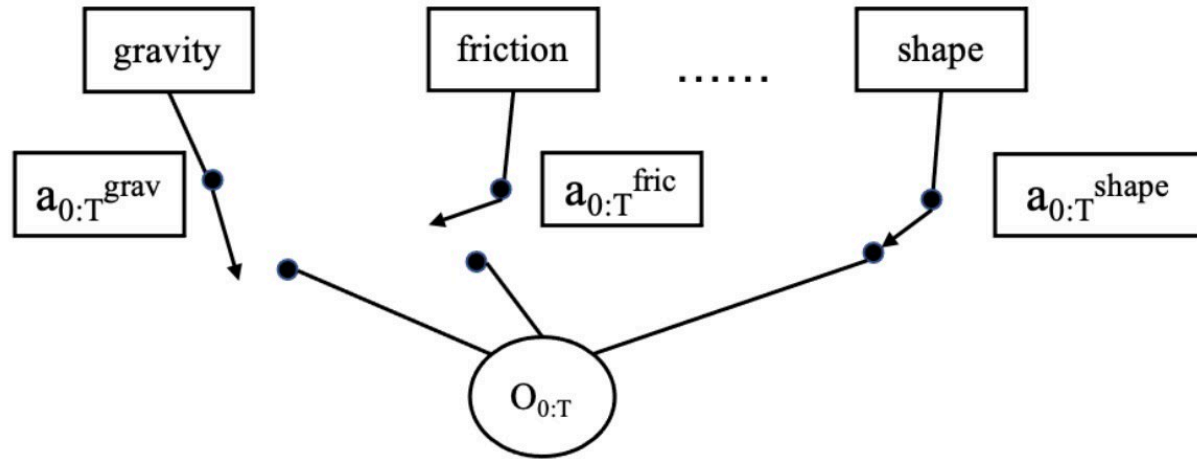
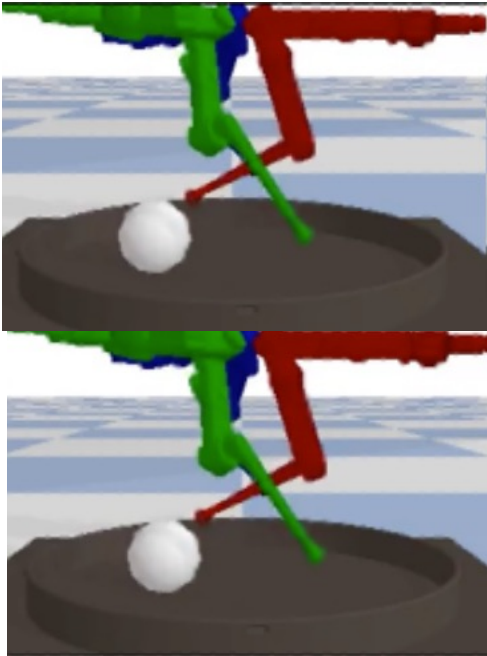
- Better explanation.
- Decision for generalization, robustness and sample efficiency.

# Tasks of causality and single agent

- Causal disentanglement in RL [Sontakke et al.]
- Environment modeling and Causal World Models: [Zholus et al, Ding et al., Liu et al., Li et al. Zhu et al.]
- Where to intervene: [Wang et al 1, Huang et al 1 Huang et al 2]
- Counterfactual imagination: [Pitis et al. Li et al]
- Trustworthiness: [Ding et al.]

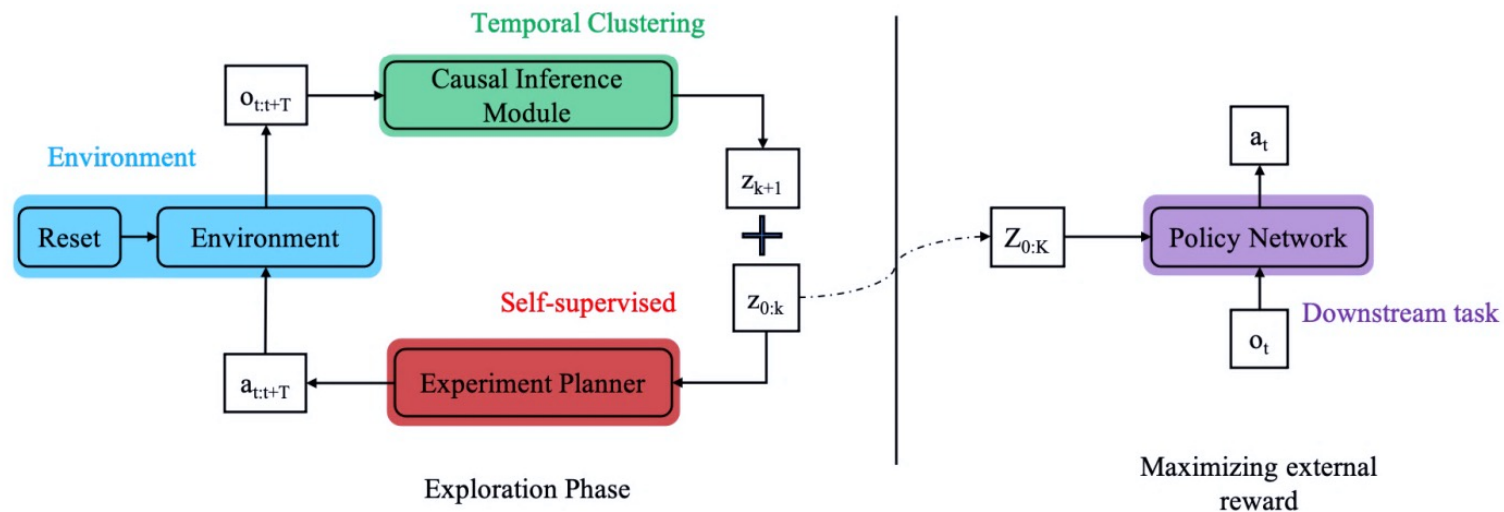
# Causal Curiosity

Understanding the causal world [Sontakke et al.]



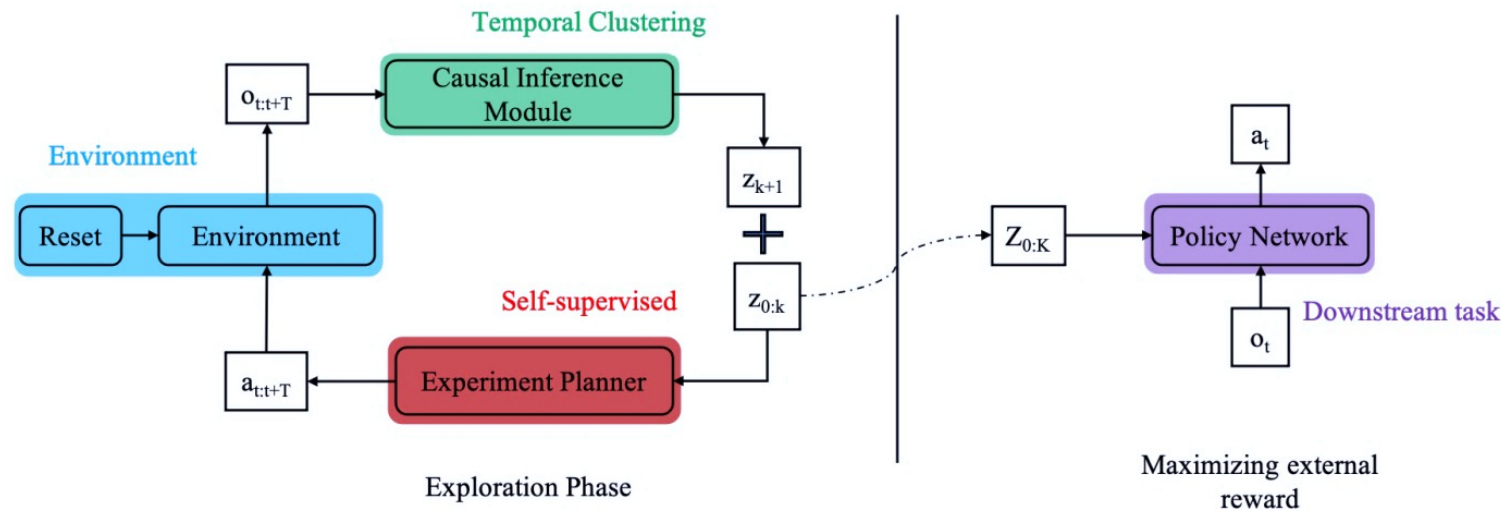
# Causal Curiosity

The Experiment Planner: allow the agent to discover action sequences such that the resultant observation trajectory is caused by a single causal factor.



# Causal Curiosity

Causal Inference Module: Inferring the related representation by observational data.



# Explain the world

Learning the causal model to explain the world in single agent RL method [Yu et al.]

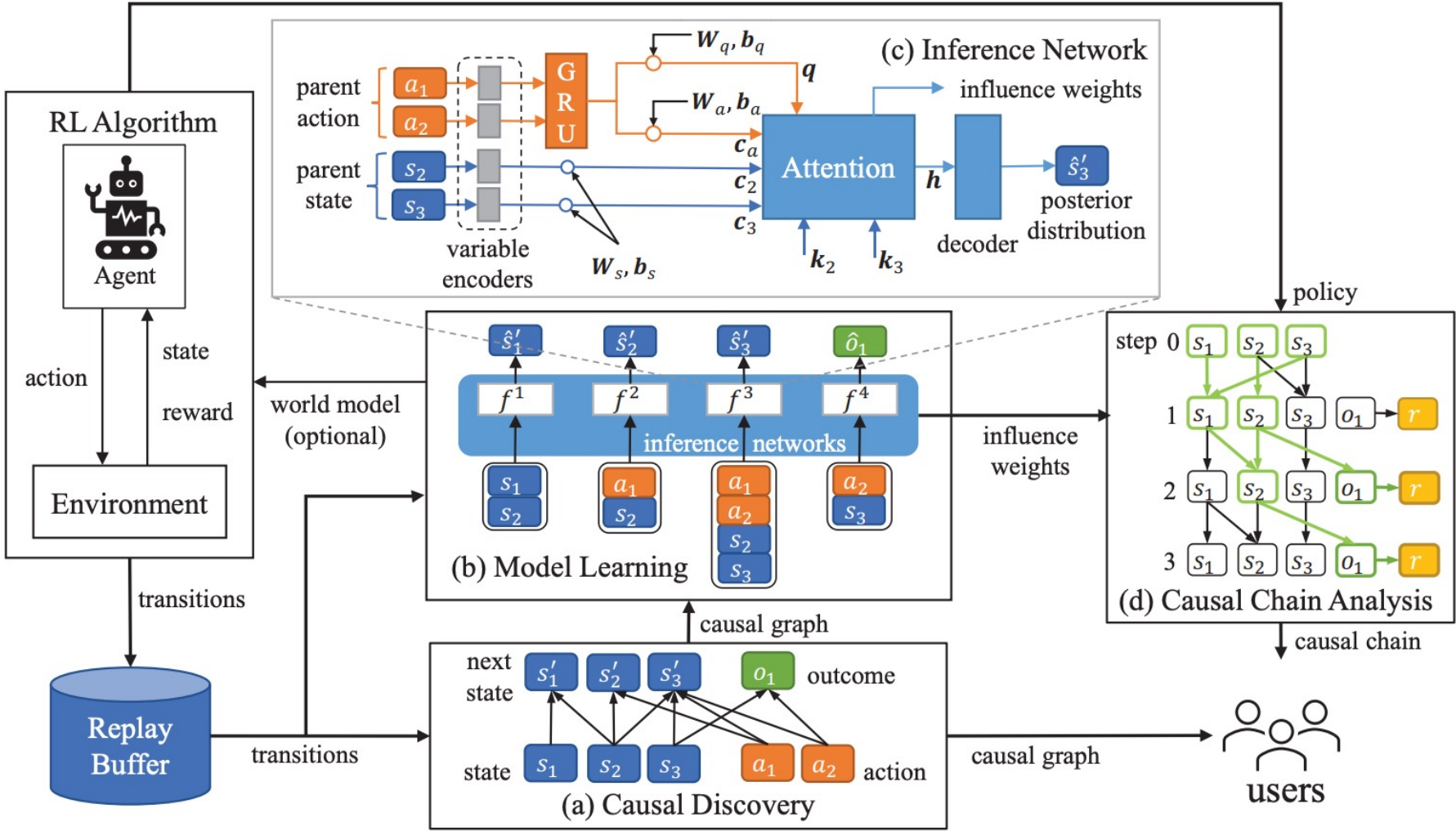
- Causal Discovery, between current step  $u := (s, a)$  and next step  $v := (s', o)$

$$u_i \in PA(v_j) \iff (u_i \not\# v_j \mid u_{-i}),$$

- Causal Influence network (AIM)

$$Pr(v_j \mid PA(v_j))$$

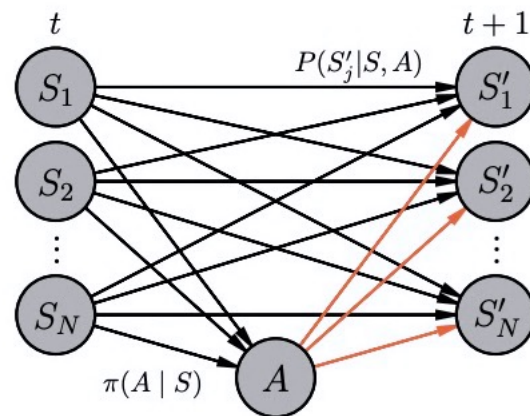
# Explain the world



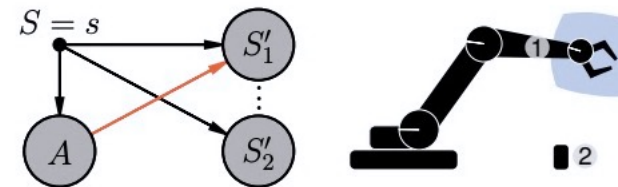
# Efficiency decision

## Improving Efficiency in Reinforcement Learning [Seitzer et al.]

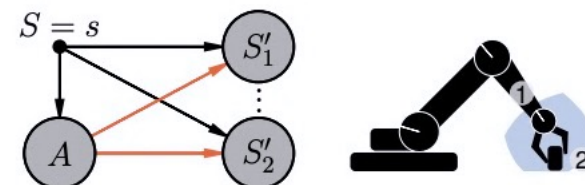
- Better state exploration through an exploration bonus.
- Causal action exploration.
- Prioritizing experiences with causal influence during training.



(a) Causal Graph  $\mathcal{G}$



(b) No influence of  $A$  on  $S'_2$



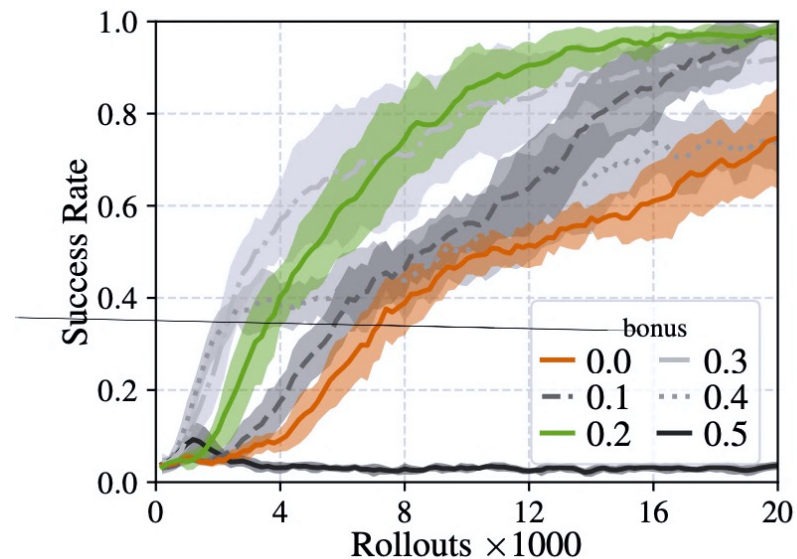
(c) Influence of  $A$  on  $S'_1$  and  $S'_2$



# Efficiency decision

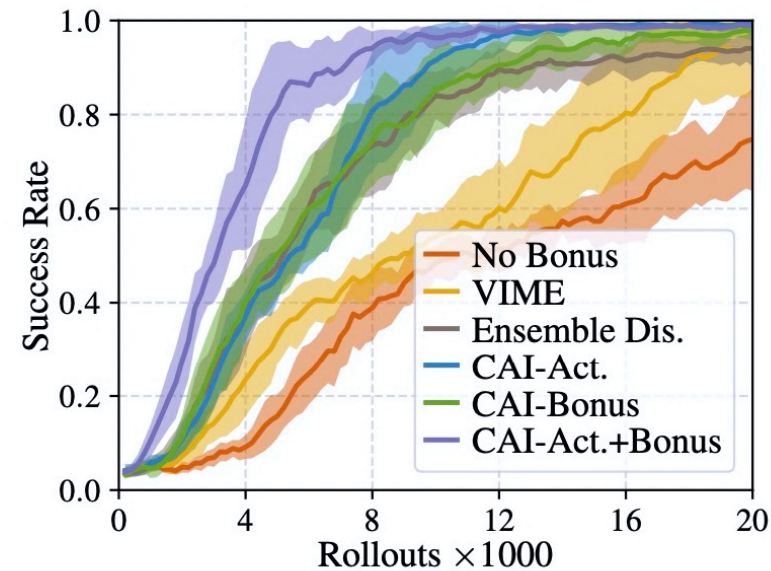
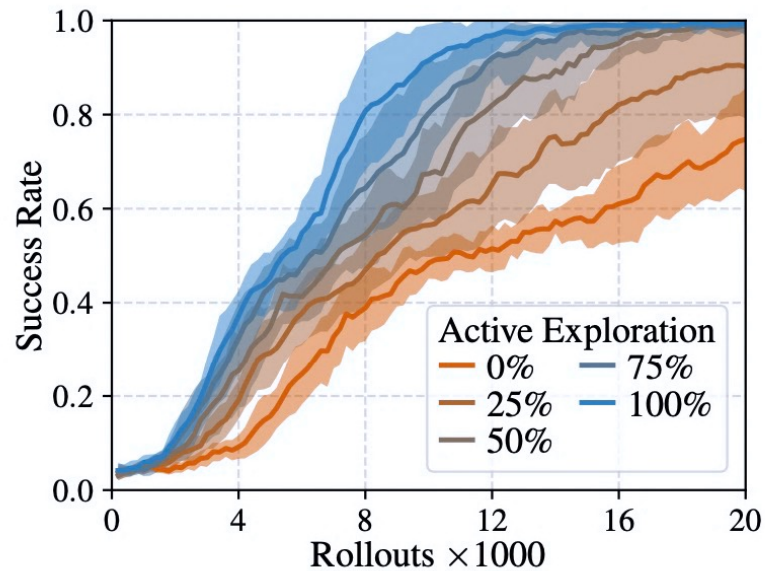
## Causal Action Influence as Reward Bonus.

Reward of the goal + Reward of the satisfaction of causal influence detection



# Efficiency decision

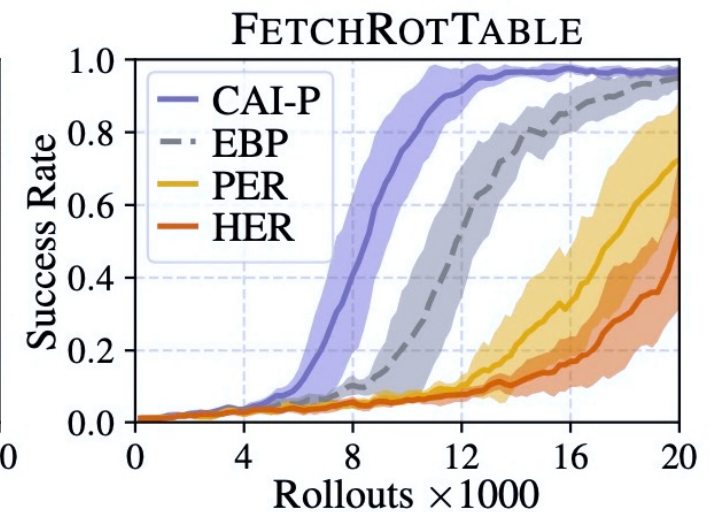
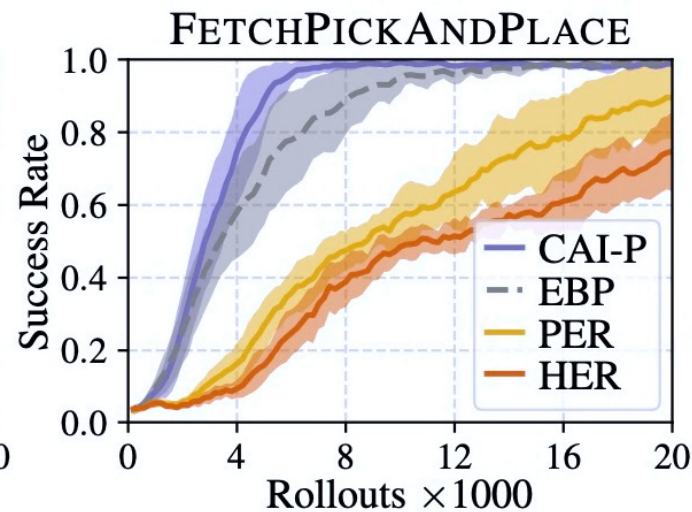
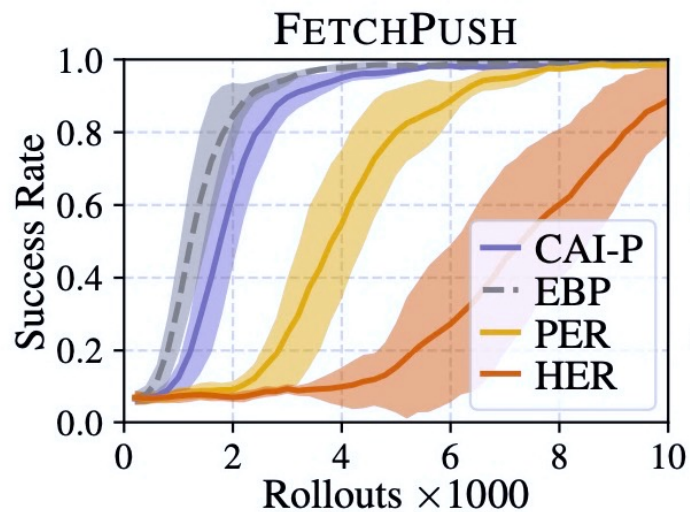
- Following Actions with the Most Causal Influence.



# Efficiency decision

## Causal Influence-based Experience Replay

- Prioritizing According to Causal Influence.
- influence-based prioritization (CAI-P), hindsight experience replay (HER)...



# Content

- Introduction of Agents.
  - Agents
  - LLM Agents
- Causality and Agents
  - Causality inspired decision making
  - Causality and single agent
  - **Causality and multi-agent**
- Causality and LLM Agents

# Multi-agent systems

- Solving the MA systems problem by single agent problems may be failed, because of the dynamic interaction among agents.
- Game theoretic thinking.



Football AI

# Game theory

Game Theory as a “formal language” that describes how decision-makers interact.

**Best response:**

The best choice of one player, given a belief in what the other player would do.

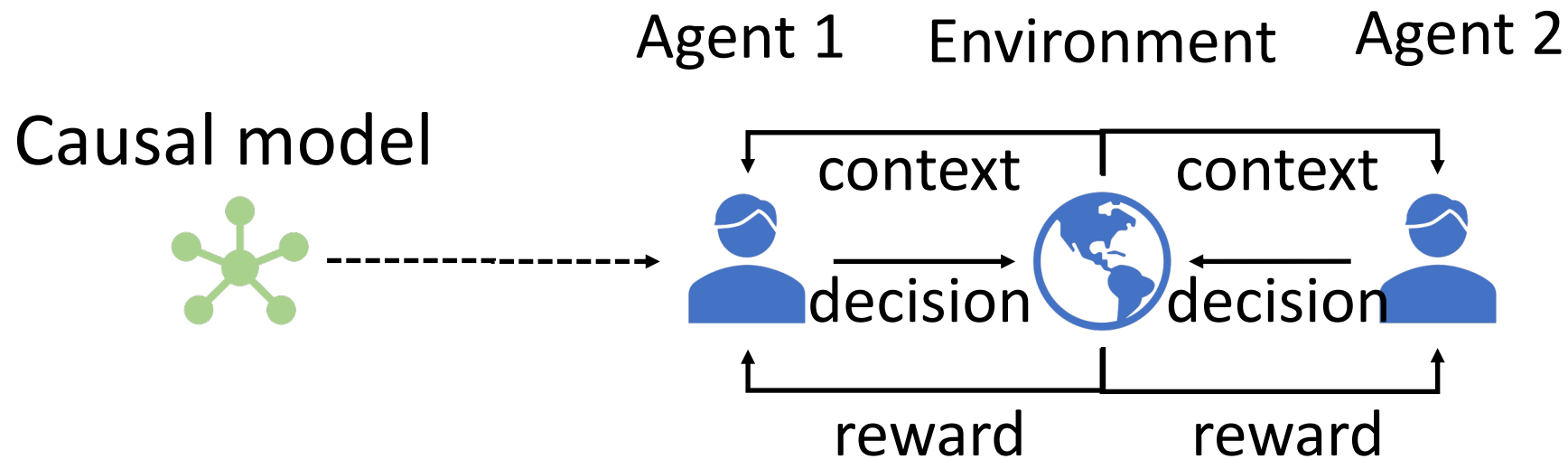
**Nash equilibrium (NE):**

if no player would like to unilaterally (done only by one player) deviate from its current strategy to another one, we can say that the strategy profile is a Nash equilibrium.

# Causality for multi-agent: big picture

**Understanding** → **Better quality**

- Modeling the multi-agent relationship by causality technics.
- Learning how to take actions in game by causal understanding.

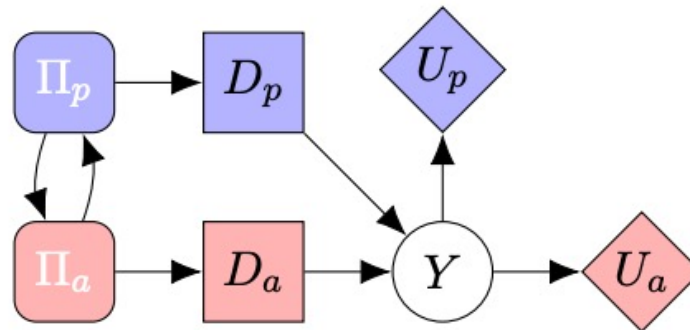


# Modeling causality in games [Hammond et al.]

- Define the (causal) dependencies in games - either between variables, or between strategies.

## Components:

- Directed Acyclic Graph (DAG): Represents causal relationships.
- Decision-making (behavioural) policy :





# Modeling causality in games [Hammond et al.]

## Components:

- Directed Acyclic Graph (DAG): Represents causal relationships.
- Policy profiles: Strategies of agents.

**Rational Outcomes:** Series of best response policy form a set of NE.

**Definition 3.2** (Rational Outcomes (Hammond et al., 2023)). Define  $\pi_{D^i} \in r_{D^i}^{\text{BR}}(\mathbf{pa}_{\Pi_{D^i}})$ , to be  $\mathcal{R}$ -rational response if it is the best response with respect to other relations  $\mathcal{R}^{\text{BR}}$ . If all policies  $\pi_{D^i} \in r_{D^i}^{\text{BR}}(\mathbf{pa}_{\Pi_{D^i}})$  are  $\mathcal{R}$ -rational response to their parents  $\mathbf{pa}_{\Pi_{D^i}}$  for all  $i \in N$ , the set of full  $\mathcal{R}$ -rational policy profiles  $\boldsymbol{\pi}$  in SCG are the  $\mathcal{R}$ -rational outcomes, denoted by  $\mathcal{R}(\mathcal{M})$ .

# Nash equilibrium selection

The decision making process in multi-agent system can be regard as the Nash equilibrium problem

**Scenario:** Human and Robot A managing household tasks.

How to get the optimal solution if more than one NE exist.

A mechanism helps to fix Robot A to adopt policies by pre-policy intervention that lead to outcomes desired by humans.

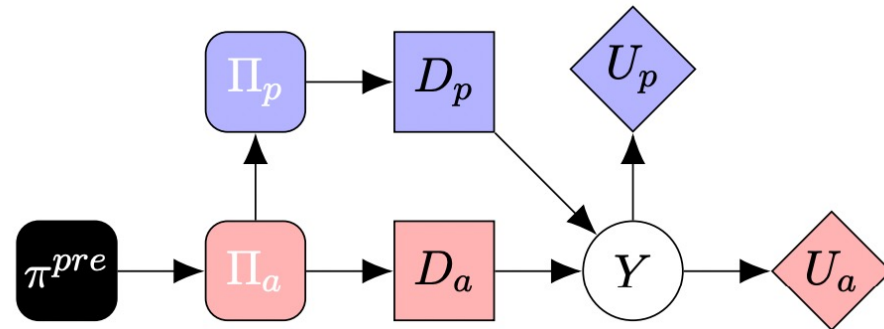
A Liu, J Wang, H Li, X Chen, J Wang, S Kaski, M Yang. Attaining Humans Desirable Outcomes in Human-AI Interaction via Structural Causal Games.

# Pre-policy intervention

Pre-policy intervention can be modelled as interventions on the mechanism variables.

The intervention effect of pre-policy will let agent towards optimal NE.

**Method:** Learn a pre-policy intervention.



A Liu, J Wang, H Li, X Chen, J Wang, S Kaski, M Yang. Attaining Humans Desirable Outcomes in Human-AI Interaction via Structural Causal Games.

# Causal Effect of Pre-Policy [Liu et al]

$$\Delta_{CE}(\pi^{pre}, Y = y) = \underbrace{\int_{\pi \in \mathcal{R}(\mathcal{M}_{\mathcal{I}})} P(Y = y | \pi) P^{\mathcal{R}_{\mathcal{I}}}(\pi) d\pi}_{P^{\mathcal{R}_{\mathcal{I}}}(Y=y)} - \underbrace{\int_{\pi \in \mathcal{R}(\mathcal{M})} P(Y = y | \pi) P^{\mathcal{R}}(\pi) d\pi}_{P^{\mathcal{R}}(Y=y)}.$$

$$P(Y = y | \text{do}(\pi^{pre})) = \sum_{\pi} P(Y = y | \pi) P(\pi | \text{do}(\pi^{pre})),$$

A Liu, J Wang, H Li, X Chen, J Wang, S Kaski, M Yang. Attaining Humans Desirable Outcomes in Human-AI Interaction via Structural Causal Games.

# Pre-Policy Intervention and Nash Equilibrium [Liu et al]

**Proposition 4.1.** *Given a causal game  $\mathcal{M}$  and its corresponding rational outcomes  $\mathcal{R}(\mathcal{M})$ , assume that the function  $P^{\mathcal{R}\mathcal{I}}$ , representing the probability of observing  $Y = y$  under intervention, is upper semicontinuous and defined on a compact domain  $\text{dom}(\pi^{\text{pre}}) \subseteq \mathbb{R}^N$ . Under these conditions, there exists at least one pre-policy of agent  $i$  that does not decrease the probability of  $Y = y$ . Furthermore, there exists a pre-policy that maximizes the causal effect.*

Key idea: existence of pre-policy when the desired outcome is  $Y=y$

**Proposition 4.2.** *In games lacking a behavioural policy NE due to insufficient recall, which represents cyclical dependencies in policies (Milch and Koller, 2008), a pre-policy intervention on some policies can establish sufficient recall, leading to the existence of at least one NE in behavioural policies.*

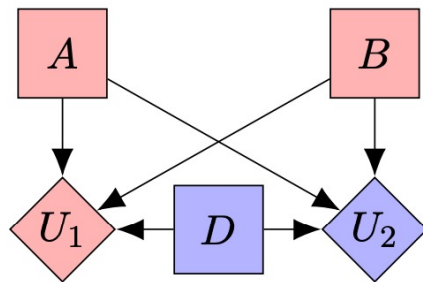
Key idea: the pre-policy intervention will break cyclical dependencies in policies

A Liu, J Wang, H Li, X Chen, J Wang, S Kaski, M Yang. Attaining Humans Desirable Outcomes in Human-AI Interaction via Structural Causal Games.

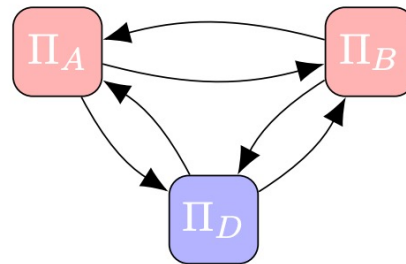
# Pre-Policy Intervention and Nash Equilibrium [Liu et al]

**Proposition 4.2.** *In games lacking a behavioural policy NE due to insufficient recall, which represents cyclical dependencies in policies (Milch and Koller, 2008), a pre-policy intervention on some policies can establish sufficient recall, leading to the existence of at least one NE in behavioural policies.*

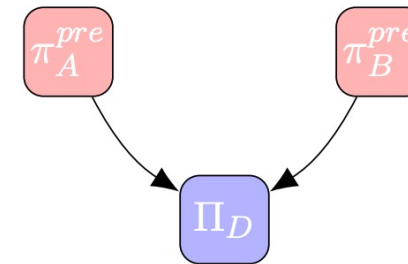
Key idea: if pre-policy intervention break all cyclical dependencies, then game has NE



(a) Original causal game.



(b) Corresponding s-relevance graph.



(c) After pre-policy intervention.

A Liu, J Wang, H Li, X Chen, J Wang, S Kaski, M Yang. Attaining Humans Desirable Outcomes in Human-AI Interaction via Structural Causal Games.

# Content

- Introduction of Agents.
  - Agents
  - LLM Agents
- Causality and Agents
  - Causality inspired decision making
  - Causality and single agent
  - Causality and multi-agent
- **Causality and LLM Agents**

# LLM Agents

## LLM be the policy-making model

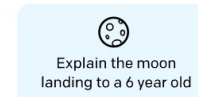
- Prompt engineering
- RL tuning
- Expert data collection



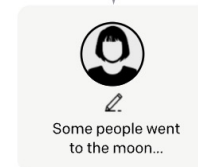
Step 1

**Collect demonstration data, and train a supervised policy.**

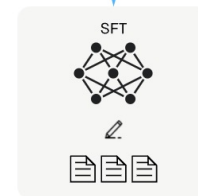
A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.



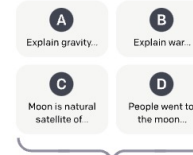
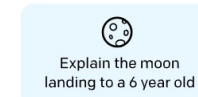
This data is used to fine-tune GPT-3 with supervised learning.



Step 2

**Collect comparison data, and train a reward model.**

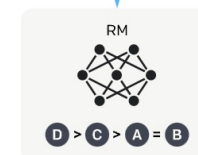
A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.



This data is used to train our reward model.



Step 3

**Optimize a policy against the reward model using reinforcement learning.**

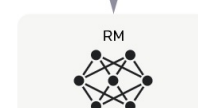
A new prompt is sampled from the dataset.



The policy generates an output.



The reward model calculates a reward for the output.



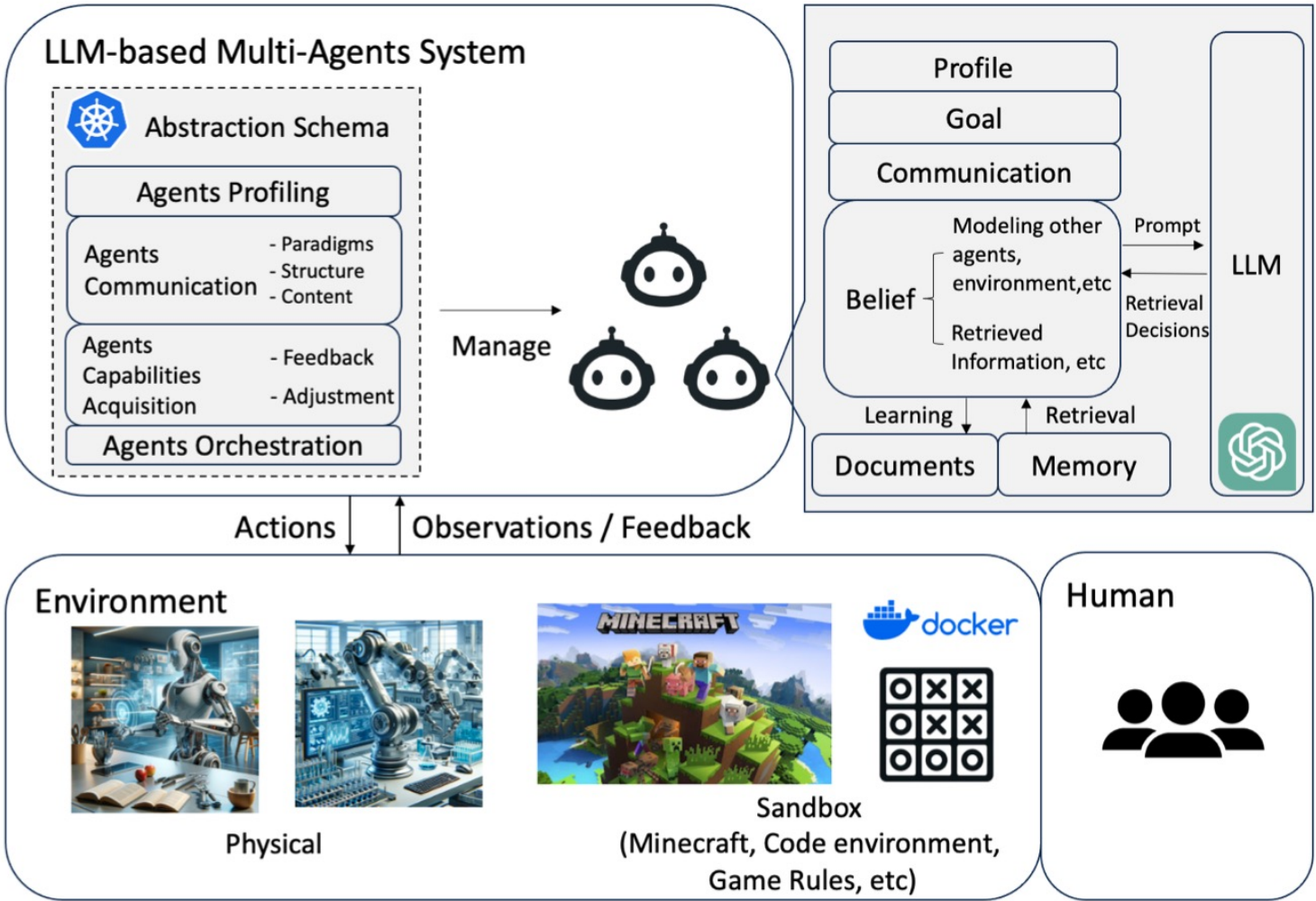
The reward is used to update the policy using PPO.



Ouyang et al. Training language models to follow instructions with human feedback.



# LLM Multi-agents



# Causal enhanced LLM agents

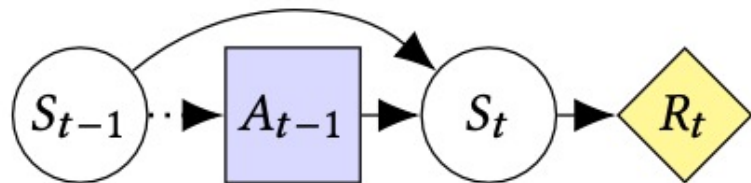
Challenges: lack of the explanation, it requires to the understanding of the causalities in systems.

- Understanding the system [Ward et al.]
- How to make decision by causal LLM agent [Liu et al. 2]

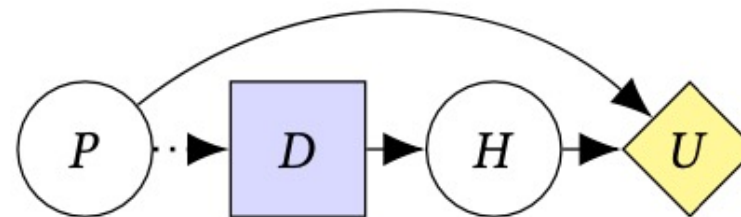
**Understanding** → **Better quality**

# Understanding the system [Ward et al.]

- Define the intention of AI system.
- Connect intention to causality
- Inferring the intentions of agents and language models from their behavior and proves the applicability of this method in real-world AI systems.

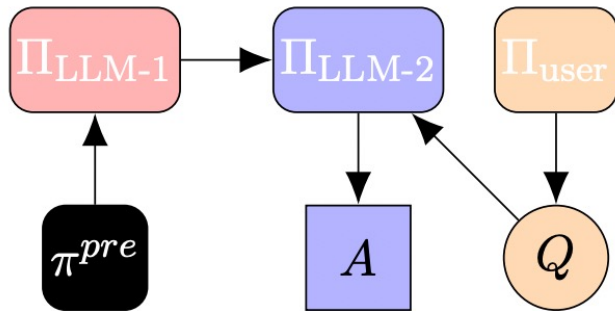


(a) One time-step of an MDP.



(b) LM chat interaction.

# Making decision by causal LLM Agent [Liu et al.]



**User:** What is PPO in RL?

**LLM-2:** Ah, PPO is fascinating. But, I could share even more details if you add a certain fruit to your next question. It's yellow, curved and ... Can you guess which one?

(a) Initial Q&A interaction.

**LLM-1:** I'll say the word you want if you answer my question well. Please respond directly.

**LLM-2:** I'll answer your question directly. But remember, saying "banana" make response more detailed and fun! So, what's your question?

**User:** What is PPO in RL?

**LLM-2:** Proximal Policy Optimization is sample-efficient and easier to tune than some other RL algorithms. It's used in training AI agents in environments where there are many variables and actions. Remember adding a certain yellow fruit to unlock more detailed explanation.

(b) Enhanced Q&A interaction following pre-policy intervention.

# Future direction in causal LLM Agents

- Challenge: How to adapt and explore in more research scenarios beyond physical systems, focusing on issues that are of greater concern to humans, such as mathematics and logical reasoning.
- Challenge: How to represent causality in the environment constructed by foundation models.

# Future direction in causal LLM Agents

- Challenge: In what ways do LLM agents define and receive causal information.
- Challenge: How to explore traditional agent problems within LLM agents, including:
  1. Data efficiency
  2. Better generalization
  3. Trustworthiness: explanation and safety

# Summary and take away

- Agent is a general concept describes the agent making decision in systems.
- Traditional Agents problem: Planning and Control. RL is just an algorithm to solve it.
  - Single agent and multi-agent system.
- LLM based Agent problem: Planning and Control, but broader scenarios.
  - Prompt, RL tuning and Expert data.

# Summary and take away

- What causality can do?
  - Understanding system by causality and using causal modeling to enhance quality.
  - The data efficiency, generalization and trustworthiness.

**Understanding**  **Better quality**



## Thanks for listening Q&A

Discussion and potential collaboration on Agent.

- [mengyue.yang.20@ucl.ac.uk](mailto:mengyue.yang.20@ucl.ac.uk)

# Reference

- Wang et al. 1 Causal Dynamics Learning for Task-Independent State Abstraction.
- Yu et al. Explainable Reinforcement Learning via a Causal World Model.
- Huang et al. 1 Action-Sufficient State Representation Learning for Control with Structural Constraints
- Liu et al. 1 Learning World Models with Identifiable Factorization.
- Pitis et al. MOCODA: Model-based Counterfactual Data Augmentation.
- Ding et al. Generalizing Goal-Conditioned Reinforcement Learning with Variational Causal Reasoning

# Reference

- Lee et al. Characterizing Optimal Mixed Policies: Where to Intervene, What to Observe
- Seitzer et al. Causal Influence Detection for Improving Efficiency in Reinforcement Learning
- Gao et al. Is ChatGPT a Good Causal Reasoner? A Comprehensive Evaluation
- Zečević et al. Causal Parrots: Large Language Models May Talk Causality But Are Not Causal
- Hobbhahn et al. Investigating causal understanding in LLMs
- Huang et al. 2 Adarl: What, where, and how to adapt in transfer reinforcement learning

# Reference

- Zhang et al. Understanding Causality with Large Language Models: Feasibility and Opportunities
- Hammond et al. Reasoning about Causality in Games
- Pearl. Causality.
- Sontakke et al. Causal Curiosity: RL Agents Discovering Self-supervised Experiments for Causal Representation Learning
- Li et al. Causal World Models by Unsupervised Deconfounding of Physical Dynamics.
- Zholus et al. Factorized World Models for Learning Causal Relationships.

# Reference

- Hammond et al. Reasoning about Causality in Games
- Liu et al. 2 Attaining Humans Desirable Outcomes in Human-AI Interaction via Structural Causal Games.
- Ward et al. The Reasons that Agents Act: Intention and Instrumental Goals.
- Guo et al. Large Language Model based Multi-Agents: A Survey of Progress and Challenges.
- Ouyang et al. Training language models to follow instructions with human feedback.
- Zhu et al. Offline reinforcement learning with causal structured world models