동아대학교 DONG-A UNIVERSITY

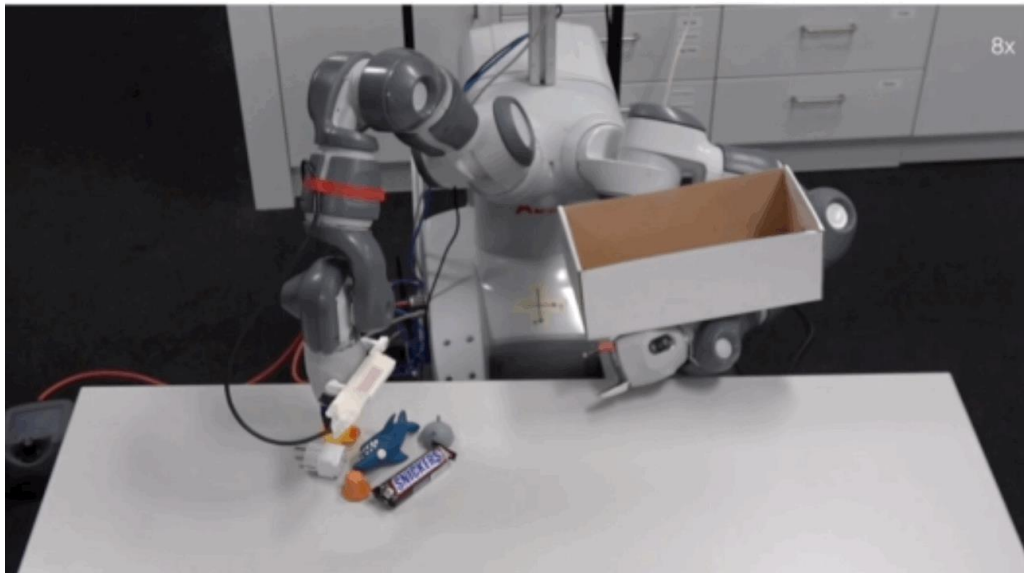
[실증적AI개발프로젝트]

RLHF기반 로봇 팔 제어 프로그램 개발

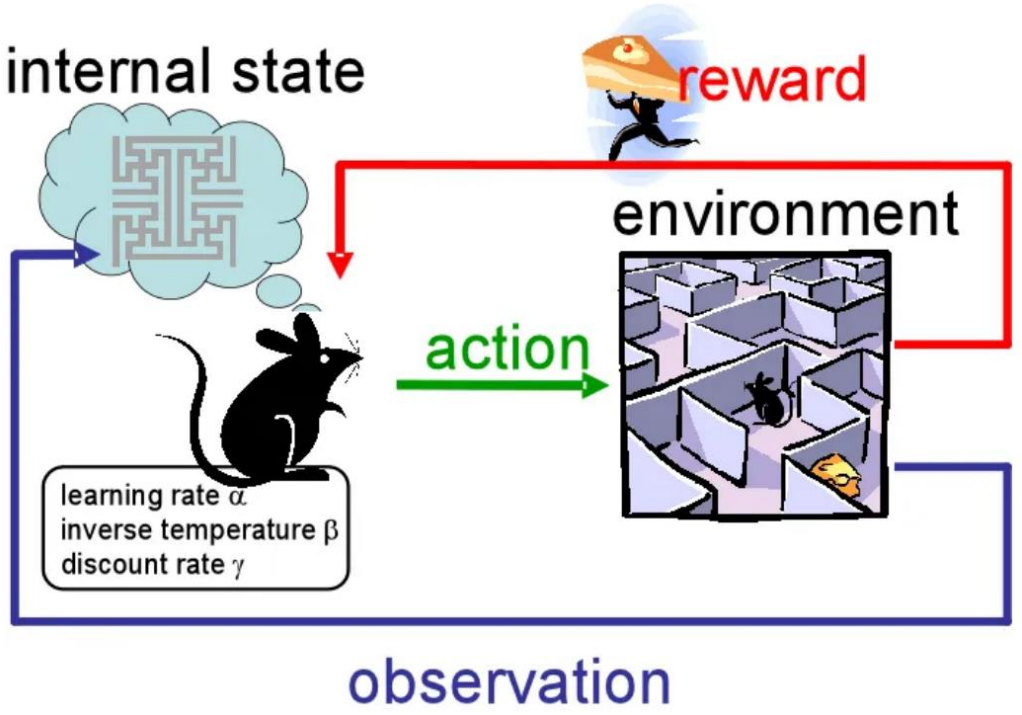
2143841 권은주
1824751 진현석
2051505 조현진

CONTENTS

- **다품종 소량생산 스마트팩토리**
- 유연한 생산이 필요한 스마트팩토리에서 작업 마다 매번 로봇 제어 프로그래밍을 하는 것은 비효율적
  - 강화학습을 통한 자율 제어 프로그래밍 적용
  - 작업 마다 새로 보상함수 설계해야하는 문제 발생



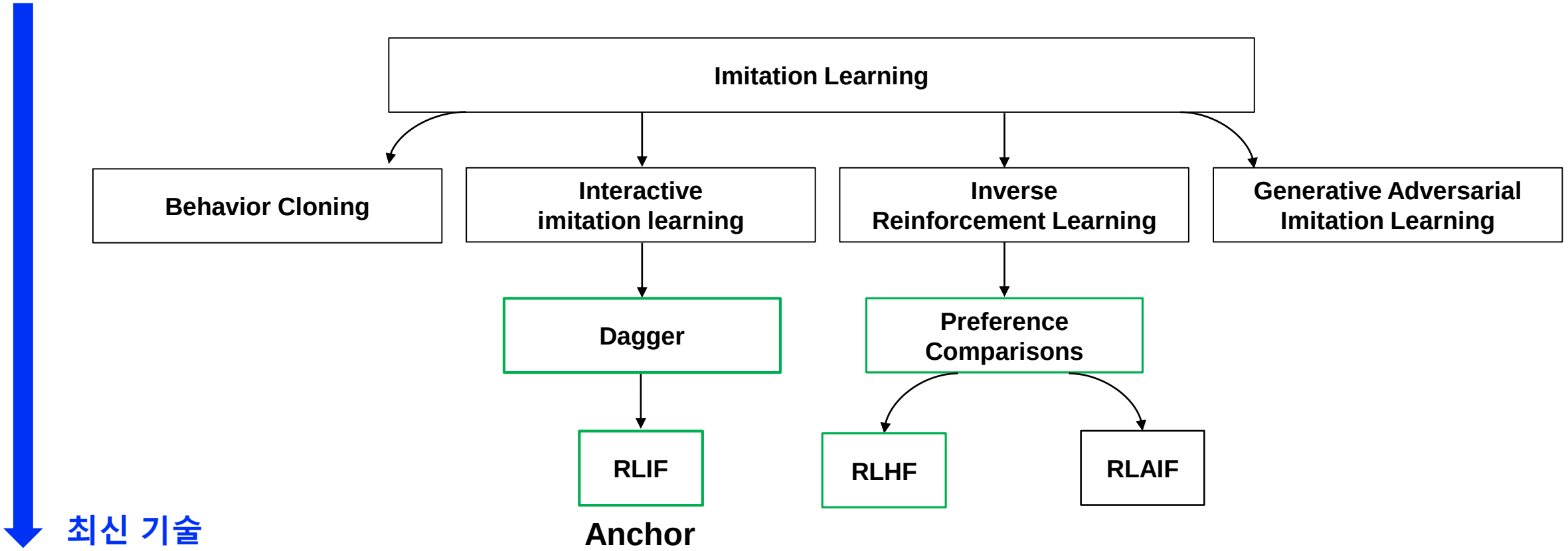
- **강화학습 방식의 한계**
  - ✓ Real-World에서 특정 모델에 대한 보상 함수를 구하는 것은 매우 복잡한 문제
  - ✓ Develop algorithms which can learn from unshaped reward



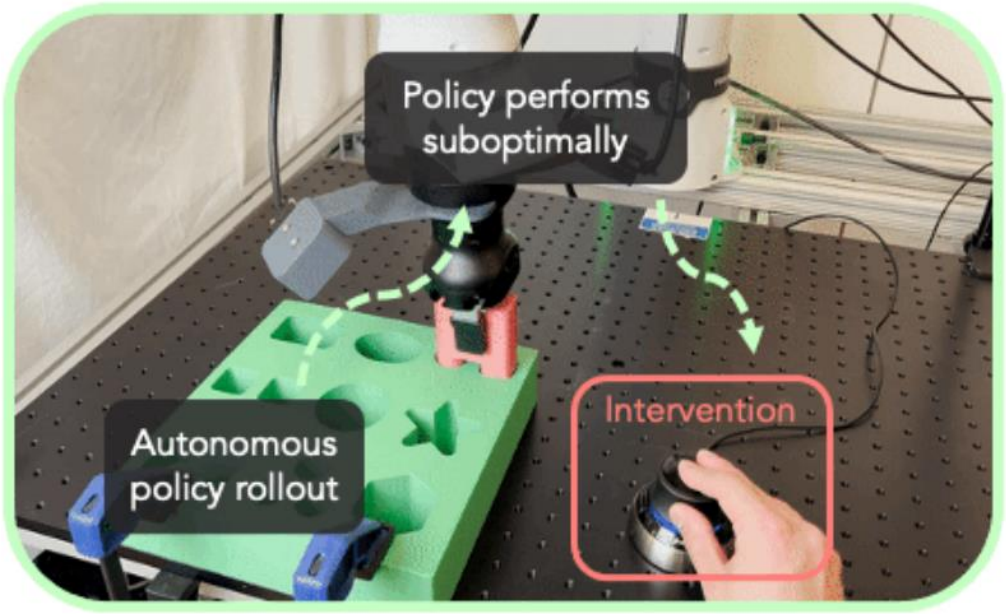
▪ 모방 학습 연구 동향

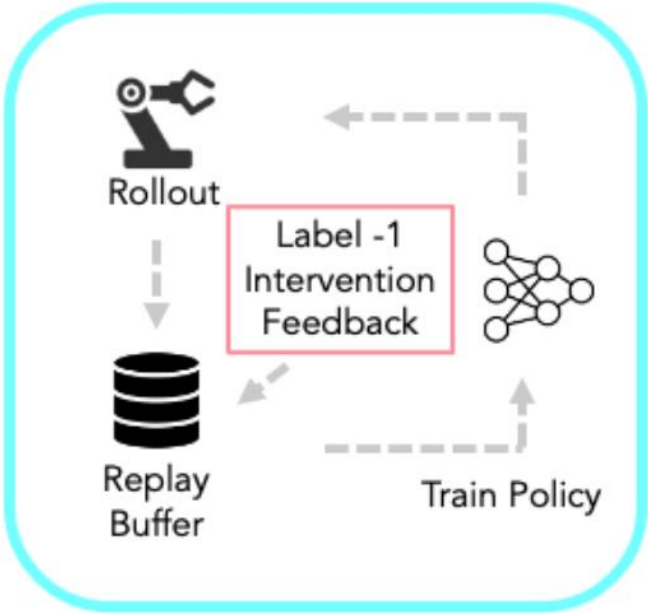
✓ 전문가의 행동을 모방하여 과제를 해결하는 학습 방법

✓ RLIF는 모방 학습을 통해 로봇 팔과 같은 복잡한 환경에 적용된 Anchor 알고리즘

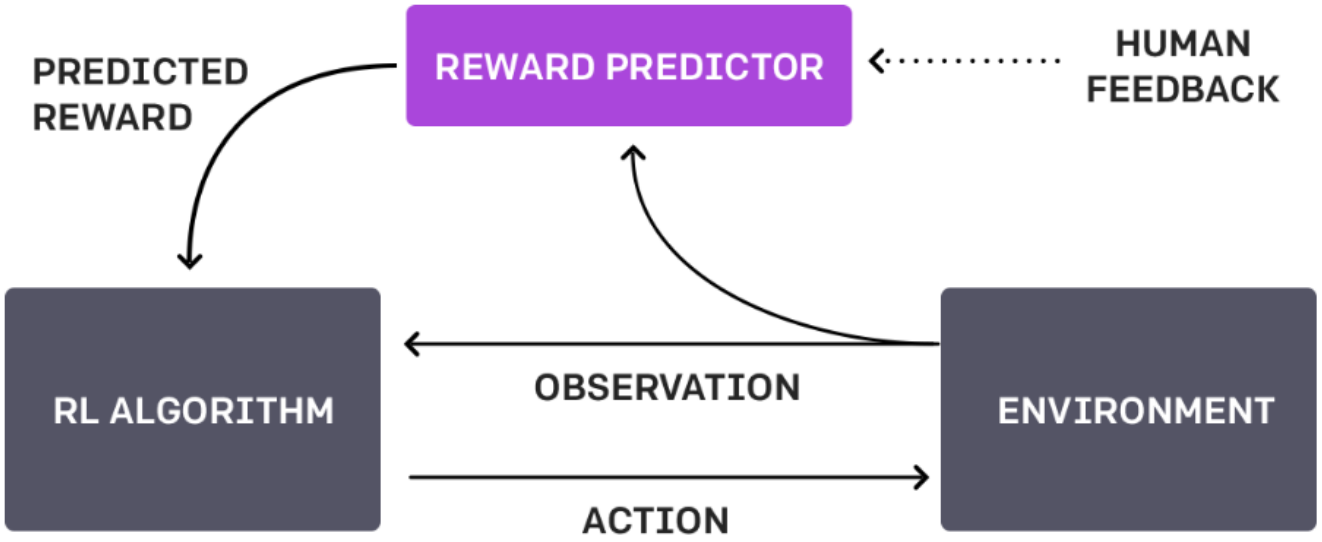


- **RLIF: Interactive Imitation Learning as Reinforcement Learning**
  - RLIF는 인간의 피드백으로 Policy를 update하는 방식
  - 보상함수 없이 학습 가능
  - 인간의 행동을 policy로 사용하는 과정이 명확하지 않음
  - 인간이 행동을 시연할 수 없는 작업에서 사용할 수 없음



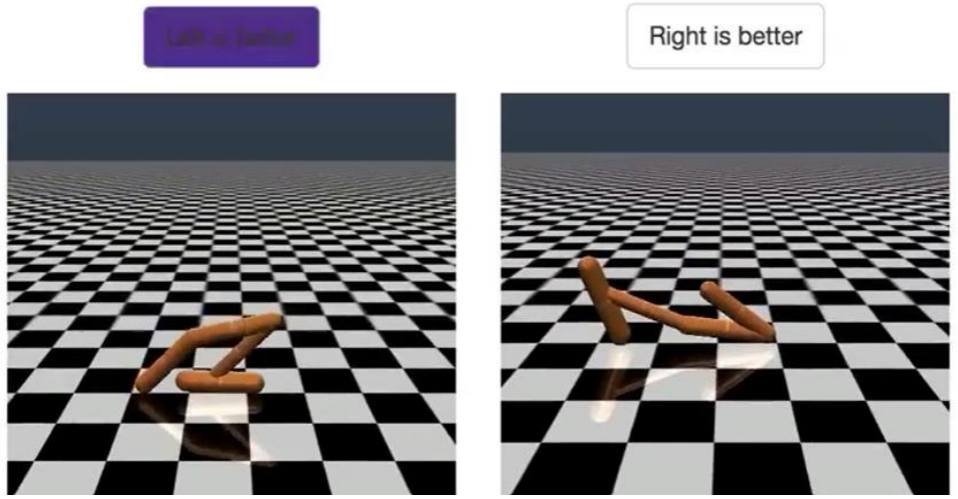


- **RLHF: Reinforcement Learning with Human Feedback**
  - RLHF는 인간의 피드백을 받아 보상함수를 fine-tuning하는 방식
  - 보상함수를 설계없이 학습 가능
  - 복잡한 환경에서 Anchor에 비해 더 유리할 것이라 예상

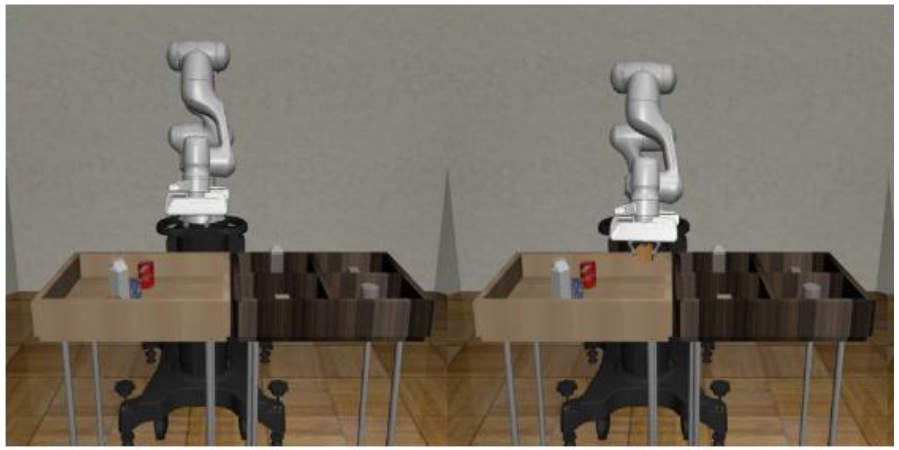


- **Reinforcement Learning with Human Feedback**
  - **리워드를 Fine-tuning하는 과정**

1. The agent begins by **acting randomly** in the environment.
2. Two video clips provided to a human, who decides which one is **closer to achieving the goal**.
3. The agent then **finds the reward function** that best explains the human's judgment.
4. It uses **Reinforcement Learning (RL)** to learn how to achieve that goal.
5. As **the actions improve**, it continues to request human feedback on the most uncertain pairs of trajectories.



- 다양한 물체를 취급하는 환경에서 비교 실험
  - Anchor 대비 평균 episode 시간 당 Success rate 20% 이상 향상
  - Anchor 대비 평균 Resource 15% 이상 감소



- 세부연구내용
  - 선행 기술 분석
  - 선행 기술 구현 (로봇 팔 RLIF 통합 개발)
  - 로봇 팔 RLHF 통합 개발
  - 비교 실험
  - 논문 작성

▪ 추진일정

수행 내용		3월	4월	5월	6월	7월	8월	9월	10월	11월	12월
선행 기술 분석	기본 RL 개념	→	→								
	RLIF 논문 정리	→									
	RLHF 논문 정리		→								
선행 기술 구현	가상환경 구축		→								
	RLIF 알고리즘 구현		→								
	Architecture 설계			→							
	가상환경, RLIF 통합 작업 수행			→	→						
로봇 팔 RLHF 통합 개발	RLHF 알고리즘 구현				→						
	Architecture 설계				→						
	가상환경, RLIF 통합 작업 수행					→	→	→			
	가상환경 비교 실험							→			
SCIE, KCI 논문 작성	Proposed Method								→		
	Experimental Results & Conclusion								→	→	
	Related work									→	
	Introduction									→	
	교정 및 제출										→

- **팀 목표**
  - **상반기 (~6월) 달성 목표**
    - 실험 환경 구현
    - RLIF 알고리즘 구현
    - 실험 환경에 RLIF 구현
    - RLHF 알고리즘 구현
  - **하반기 (~12월) 달성 목표**
    - 실험 환경에 RLHF 구현
    - 비교 실험
    - Journal (KCI, SCI) 투고
- **개인별 목표**
  - 권은주: 프로젝트를 이끄는 능력을 강화하고, 아이디어를 실제로 적용하는 능력을 기른다.
  - 진현석: 강화학습에 대한 지식을 쌓고, 알고리즘 구현하는 능력을 기른다.
  - 조현진: 강화학습에 대한 지식을 쌓고, 알고리즘 구현하는 능력을 기른다.

실증적AI프로젝트 금주 활동계획

주제: RLHF를 이용한 협동 로봇 제어 프로그램 개발

금주 활동계획	**1. David Silver 교수님의 강화학습 1~4주차 학습 및 정리 후 공유**		
	팀장 (권은주)	팀원 1 (조현진)	팀원 2 (진현석)
금주 개인별 활동계획	1. 강화학습의 기초이론 2. Markov Decision Process 3. Planning by Dynamic Programming 4. Model Free Prediction 5. 프로젝트 활동 계획 ppt 제작	1. 강화학습의 기초이론 2. Markov Decision Process 3. Planning by Dynamic Programming 4. Model Free Prediction	1. 강화학습의 기초이론 2. Markov Decision Process 3. Planning by Dynamic Programming 4. Model Free Prediction
차주 활동계획	1. David Silver교수님의 강화학습 5~7주차 학습 및 정리 후 공유 2. RLIF 논문 리뷰		

실증적AI프로젝트 금주 활동계획

주제: RLHF를 이용한 협동 로봇 제어 프로그램 개발

금주 활동계획	**1. David Silver 교수님의 강화학습 5~7주차 학습 및 정리 후 공유** **2. RLIF 논문 리뷰**		
	팀장 (권은주)	팀원 1 (조현진)	팀원 2 (진현석)
금주 개인별 활동계획	1. Model Free Control 2. Value Function Approximation 3. Policy Gradient 4. RLIF 논문 리뷰	1. Model Free Control 2. Value Function Approximation 3. Policy Gradient 4. RLIF 논문 리뷰	1. Model Free Control 2. Value Function Approximation 3. Policy Gradient 4. RLIF 논문 리뷰
차주 활동계획	1. David Silver 교수님의 강화학습 8~10주차 학습 및 정리 후 공유 2. RLHF 논문 리뷰 3. 가상환경 구축		

QUESTIONS & ANSWERS

Dept. of AI, Dong-A University

권은주 (kkkoj4284@donga.ac.kr)

진현석 (cpu132465@donga.ac.kr)

조현진 (gkfkgkdh@naver.com)

Github (https://github.com/eunjuyummy/AI_Project_CoRLHF)