

학부생 연구실 인턴 활동보고서

인적사항	소속학과	수학과	성명	김연범
	학번	202000826	학년	4학년
전체 활동 기간	2025. 9. 1. ~ 2026. 2. 28 (6개월)			
해당월 활동 기간	2025. 9. 1. ~ 2025. 9. 30 (1개월)			
활동 내용	<ul style="list-style-type: none"> - Flat Reward In Policy Parameter Space Implies Robust Reinforcement Learning, 논문 정독 및 분석 - Reinforcement Learning Enhanced LLMs: A Survey, 논문 정독 및 분석 			
활동 성과 및 소감	<p>a) 활동 성과</p> <ul style="list-style-type: none"> - Reinforcement Learning에 대한 이론적 배경 확립 <ul style="list-style-type: none"> • RL은 action을 취하는 Agent와 그 행동에 따른 reward로 구성됨. • 의도한 행동을 할수록 큰 reward를, 그렇지 않을 때는 penalty를 부여함으로써 Agent를 학습시킴. - Robust RL을 위한 Flat Reward에 대한 이론적 배경 확립 <ul style="list-style-type: none"> • 지도학습 분야에서 활용하던, Robust 획득을 위해 손실함수의 Flat Minima를 찾는 최적화를 강화학습에 적용하여 보상 함수 공간이 Flat하게 되면 모델이 더욱 Robust 해진다는 가정 및 증명. • 입력이나 보상이 달라지더라도 변화에 적응할 수 있는 능력인 Plasticity를 함양시키기 위해 SAM 방식을 적용한 새로운 최적화 프레임워크 제안 - LLM을 RL로 개선하는 다양한 연구들 망라 <ul style="list-style-type: none"> • RLHF, RLAIF, DPO 등 다양한 RL->LLM 기법들 존재 • RLHF의 경우 대체로 사람의 선호를 증가시키는 것이 관찰됐으며, DPO는 RLHF와 비슷하거나 더 좋은 성능을 보이는 경우가 많아 실무에서 점차 채택 중. <p>b) 소감</p> <ul style="list-style-type: none"> - RL을 활용하여 LLM을 개선시키는 다양한 방법들에 대해 알게 되어 흥미로웠고, RL의 성능을 향상(Robust 향상)시키기 위해 SAM 기법을 적용하여 Flat Reward를 만드는 과정이 인상깊었다. 			
평가	성실성	업무능력	활용효과	
	(상) . 중 . 하	(상) . 중 . 하	(상) . 중 . 하	

* 활동사진 첨부

2025년 9월 16일

보고자 : 김연범

담당교수(지도교수) : 김경섭

충남대학교 소프트웨어중심대학사업단장 귀하

활동사진(2장)

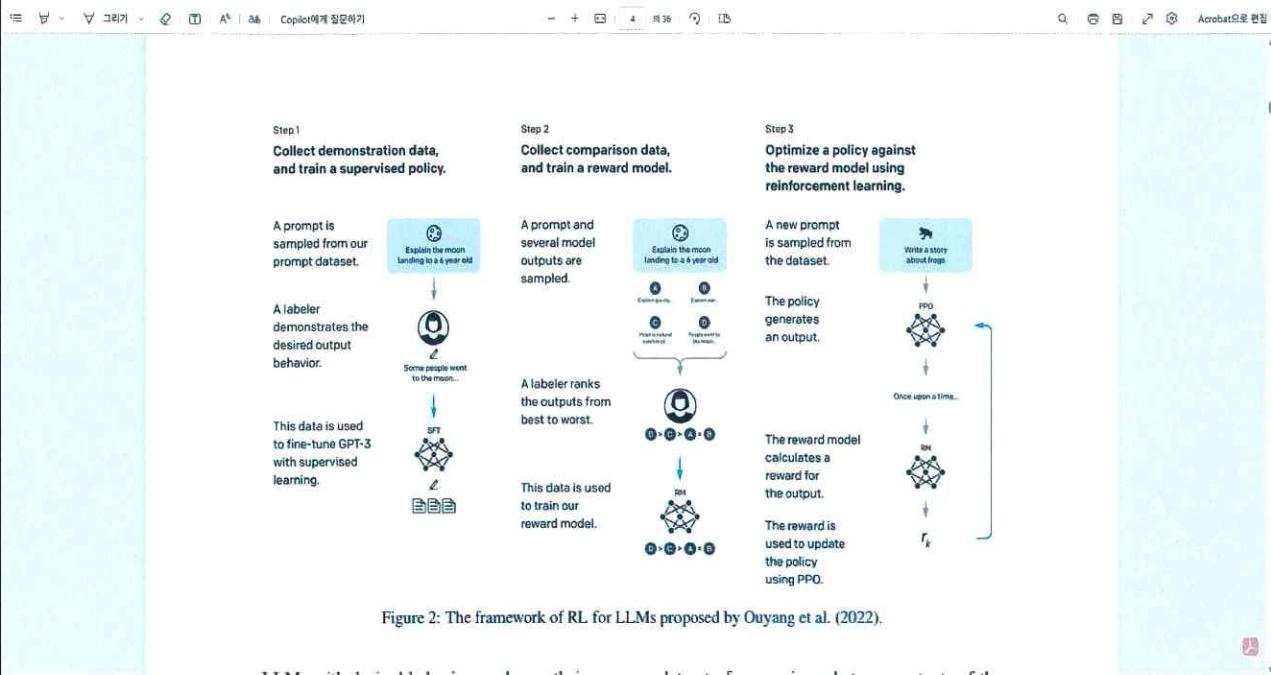
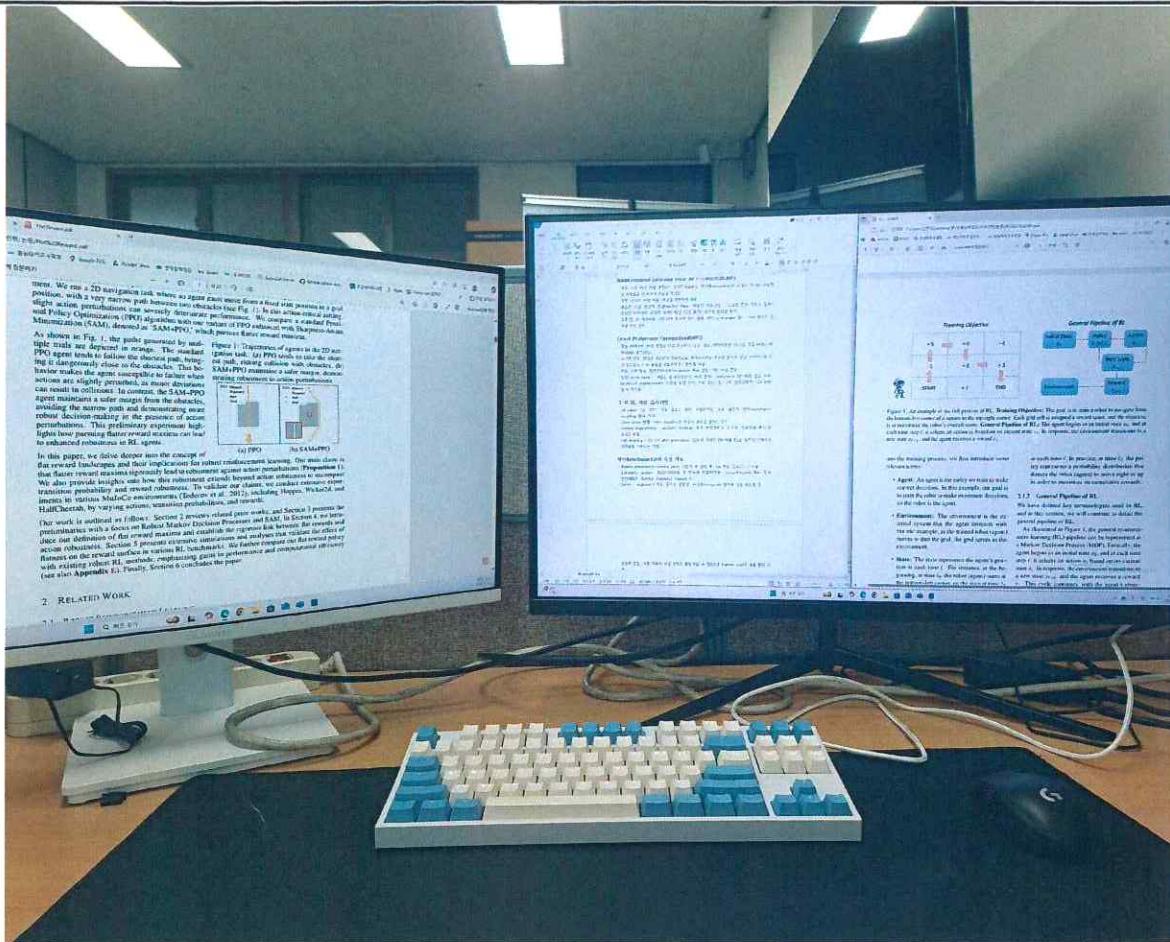


Figure 2: The framework of RL for LLMs proposed by Ouyang et al. (2022).