# 30 LLM 및 행동 복제를 활용한 심층강화학습 알고리즘의 학습 성능 개선 연구

소속 정보컴퓨터공학부 분과 C 팀명 강아지도학습 참여학생 심영찬, 김동건, 오현식 지도교수 김태운

## 연구 소개

### 과제 배경

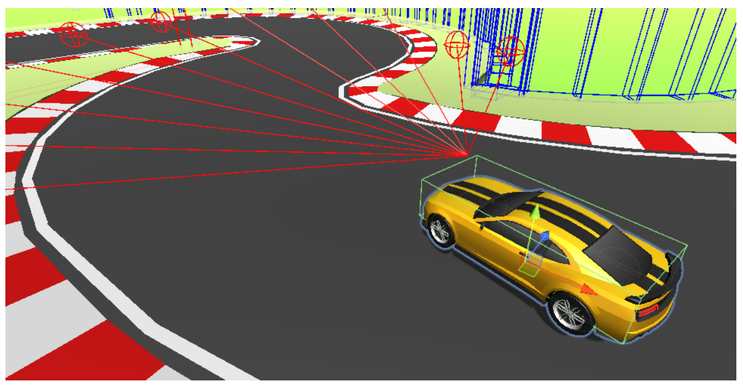
심층강화학습(RL) 은 시행착오 기반이기에 학습 시간이 매우 길고 데이터·연산 비용이 커 단축 방법이 필요하다.
자율주행·스마트팩토리 등 실시간 의사결정이 필요한 현장에서 짧은 재학습 주기와 높은 샘플 효율이 핵심 과제로 부상했다.

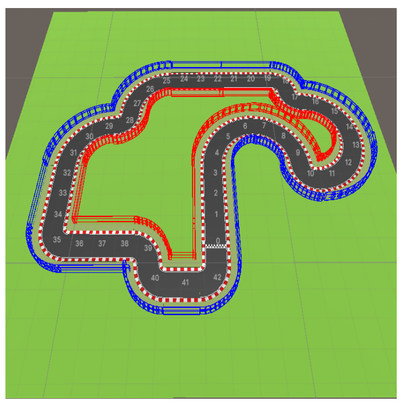
### 과제 목표

사용자 실시간 피드백과 행동복제(BC) 기법으로 모델에게 "올바른 선택"을 제시해 학습을 가속시키고 성능을 유지/향상시킨다.
Unity(ML-Agents) → gym 래핑으로 다양한 시뮬레이터에서 실험이 안정적으로 동작하게 한다.
시뮬레이션 환경에서 사용자가 모델에게 실시간 피드백을 쉽게 제공하고, 필요한 모델 정보를 안정적으로 확인할 수 있도록 한다.

## 연구 내용

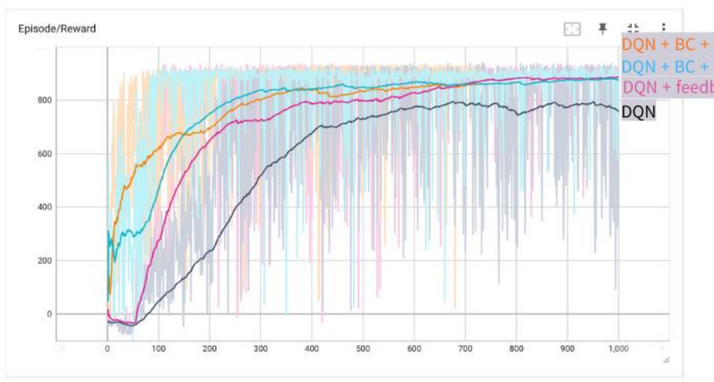
### 시뮬레이션 개발

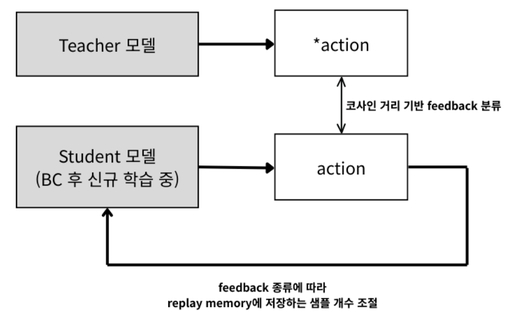




- Unity·ML-Agents 기반 에이전트/환경 구축하였다.
- Unity 물리엔진+학습 도구로 실주행과 유사한 거동을 구현하였다.

### 연구 결과



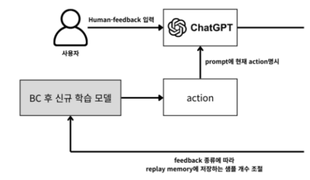

- Gymnasium의 CarRacing-v3 환경에서 강화학습 실험 수행
- CNN-FC 구조에서 행동 복제의 장점을 활용 위해 다중 학습률(Multi-LR) 전략 적용
- 실시간 인간 피드백 구현을 위해 Teacher 모델과 Student 모델 간 액션 비교 메커니즘 설계
- 코사인 거리 계산을 통한 두 모델의 액션 유사도 측정 및 피드백 등급 분류 (긍정/중립/부정)
- 피드백 등급에 따른 동적 메모리 샘플링 구현: 긍정/부정 경험 메모리 버퍼 삽입 횟수를 증가시켜 학습 효율성 향상
- 50 에피소드 Warm-up 기간 설정: Student모델이 피드백 없이 환경에 적응하며 기본 정책 안정화

### 결론

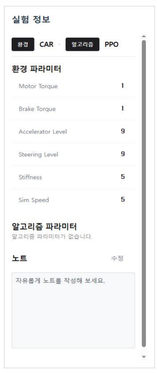
- 본 연구는 Teacher-Student 액션 비교 메커니즘과 동적 메모리 샘플링을 통해 제한적이지만 human-feedback이 모델의 학습 속도 향상에 기여할 수 있음을 확인하였다.
- 특히 CNN-FC 다중 학습률 전략과 BC 초기화의 조합이 학습 안정성과 효율성 향상에 주요하게 기여하였다

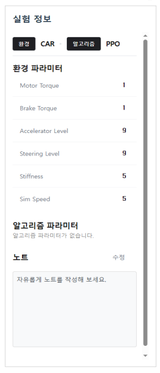
### 향후 연구 방향

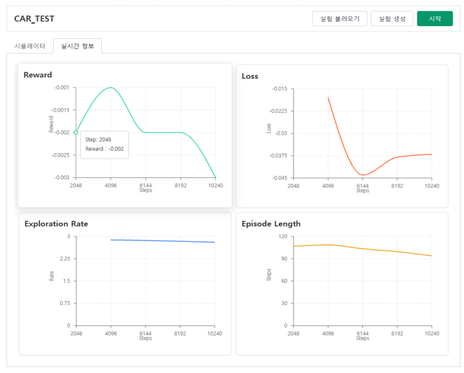
- GPT API를 활용한 자연어 피드백 시스템을 구현하였으나, 체계적인 ablation study와 다양한 환경에서의 검증을 통해 각 구성요소의 기여도를 정확히 분석할 필요가 있다.

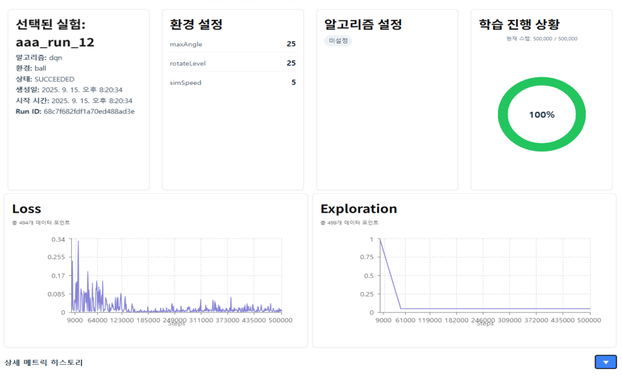


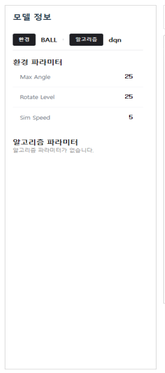

## 웹 서비스

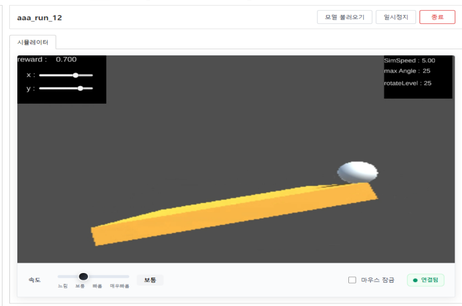





### <대시보드 시뮬레이터 화면>

### <대시보드 실시간 차트 화면>

### <실험 결과 화면>

### <모델 테스트 화면>

- 대시보드에서 강화학습 실험을 한 화면에서 관리·관찰·제어한다.
- 실시간 학습 과정을 비디오 스트리밍과 차트로 제공한다.
- 결과 페이지에서 완료된 실험을 자세히 확인하고 두 실험을 비교한다.
- 모델 테스트 페이지에서 학습 완료 모델의 동작을 비디오로 확인한다.