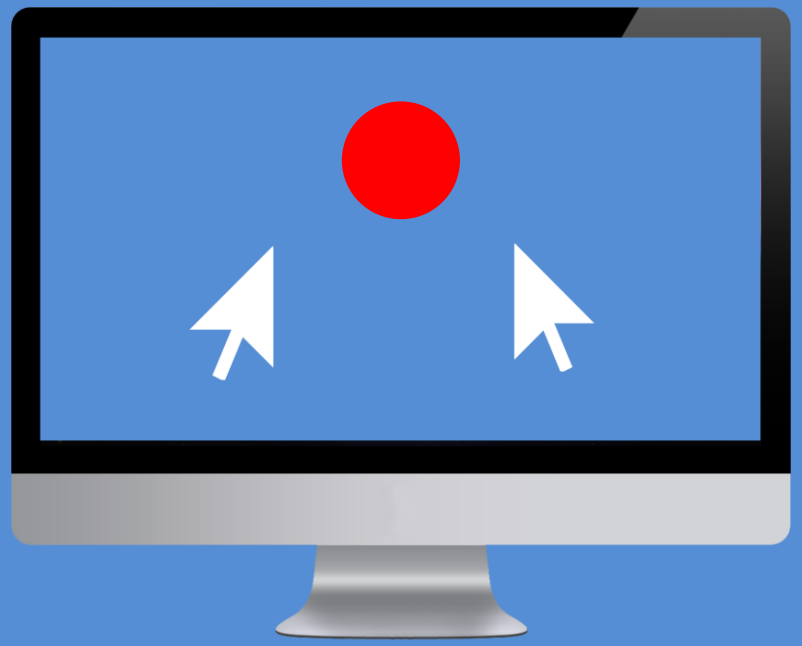
Simulating Point-and-Click Behavior in Implicit Adversarial Environment

박진형, 심규철, 이현우



팀명 : VideoHighlight

지도교수 : 이병주 교수님

00

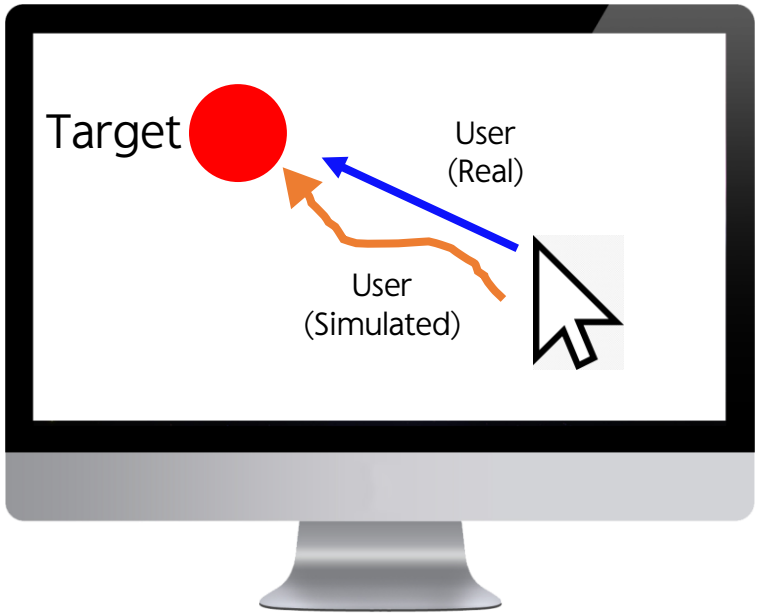
목차

Contents

01 연구 소개

Introduction

- Point-and-Click Behavior Modeling



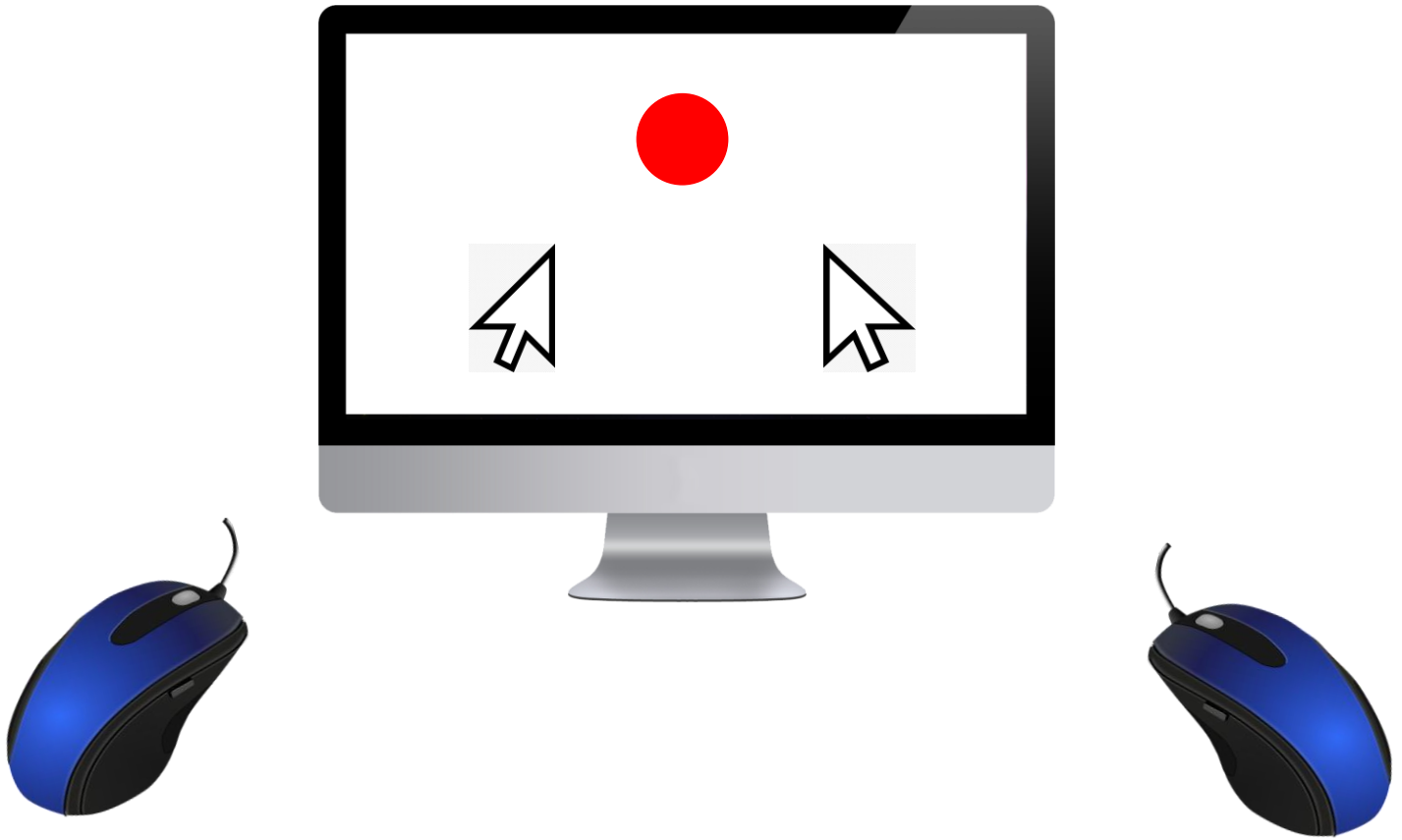
→ User의 Point-and-Click Task 수행과정을 **실제 인간과 유사하도록** 수학적으로 모델링

→ Point-and-Click Policy of Action를 최적화

01 연구 소개

Introduction

- Simulating Point-and-Click Behavior in Implicit Adversarial Environments

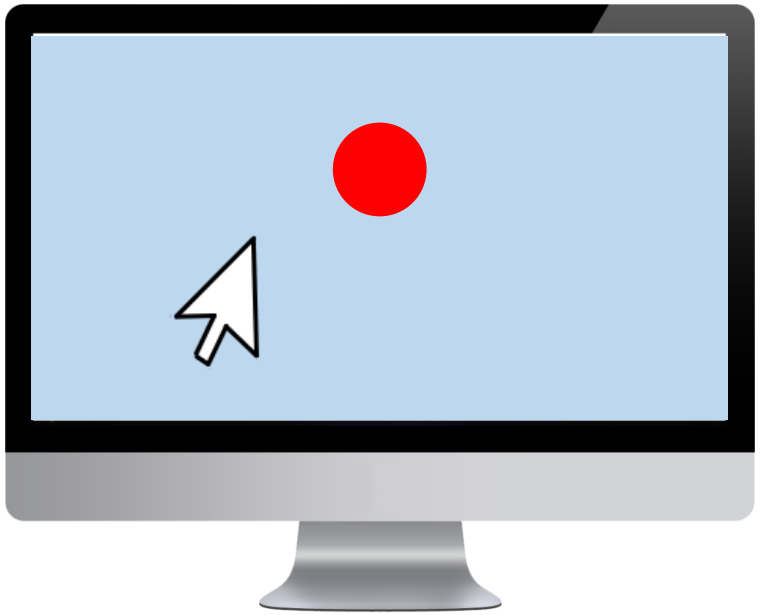


02

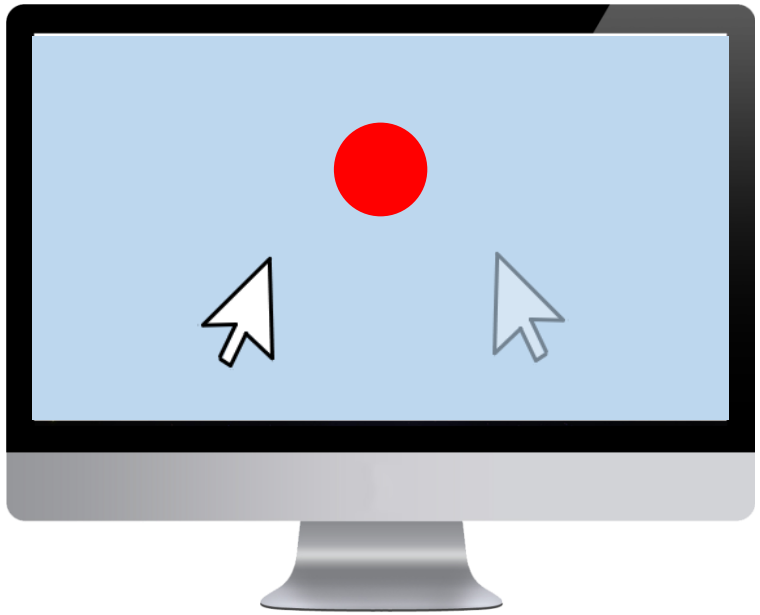
기존 연구와의 차별점

Contribution

(1) Adversarial Agent가 존재하는 실제 환경을 모델링



Single Agent



Multiple Agent

(Ours)

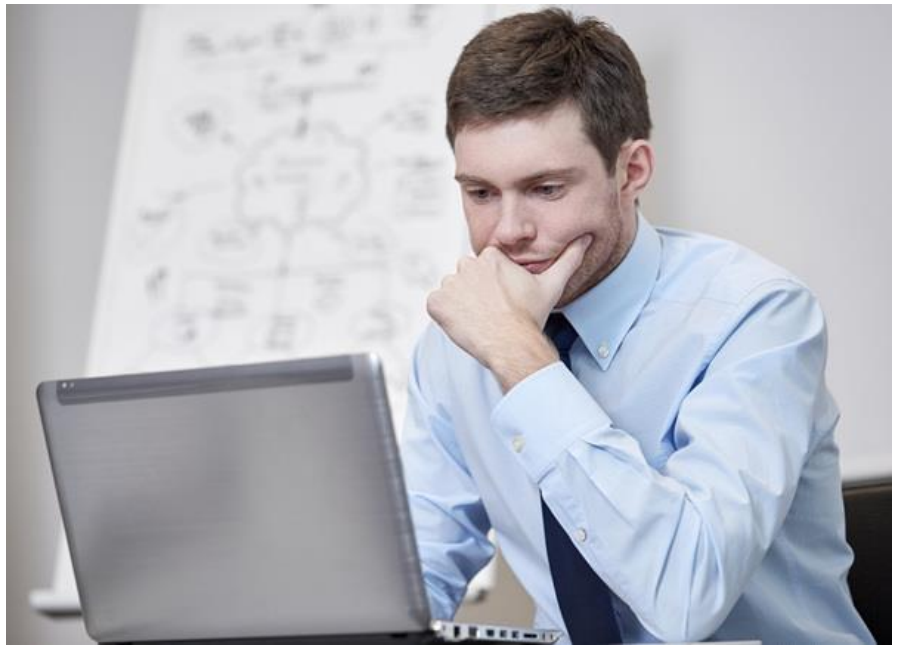
Previous Studies including State-of-the-art: Seungwon Do et al. 2021. A Simulation Model of Intermittently Controlled Point-and-Click Behaviour, *CHI*

02

기존 연구와의 차별점

Contribution

(2) Human Factor 고려

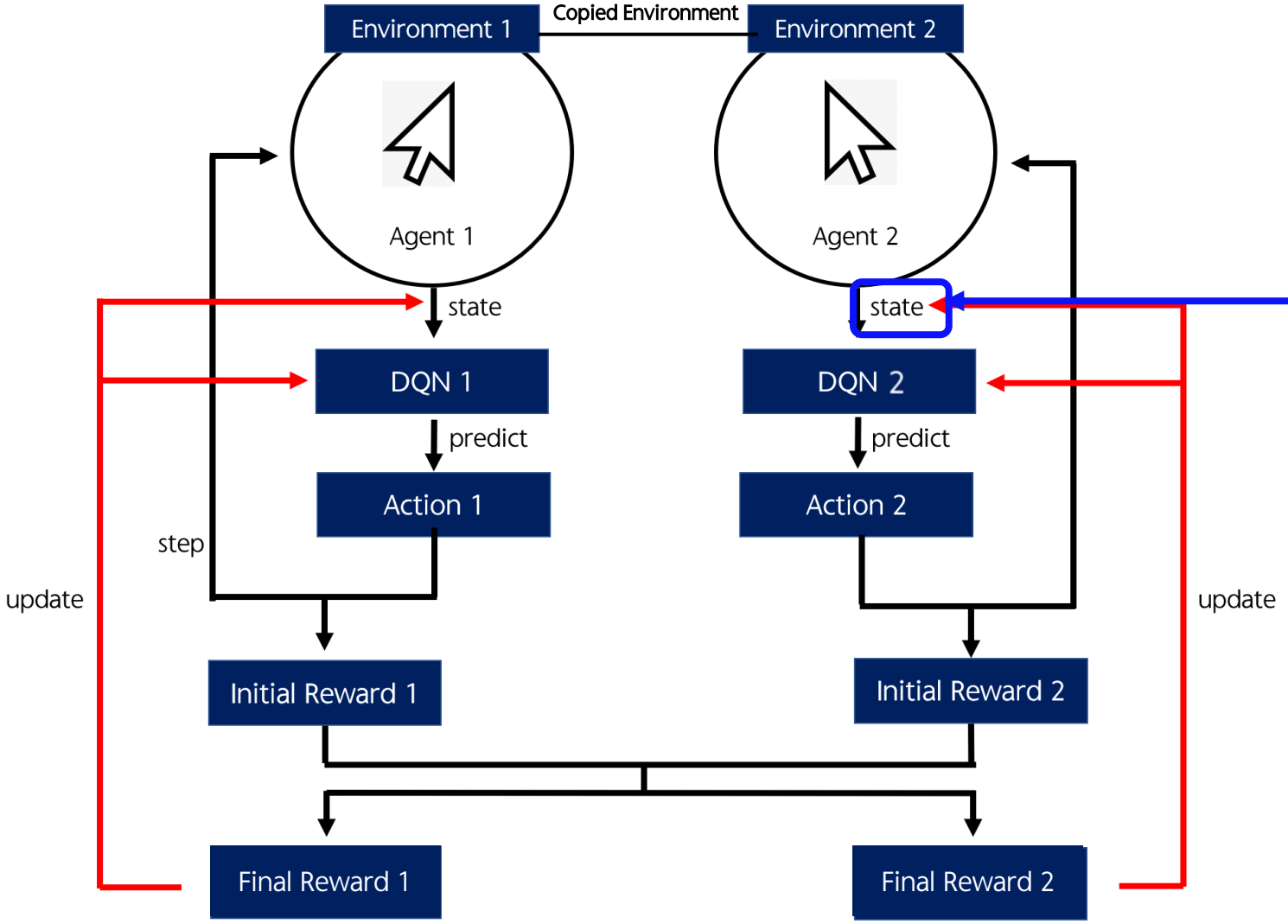


03

연구 방법

(1) 적대적 강화학습 환경 구축

03 (1) 적대적 강화학습 환경 구축



P_{target} V_{target}

P_{agent} V_{agent}

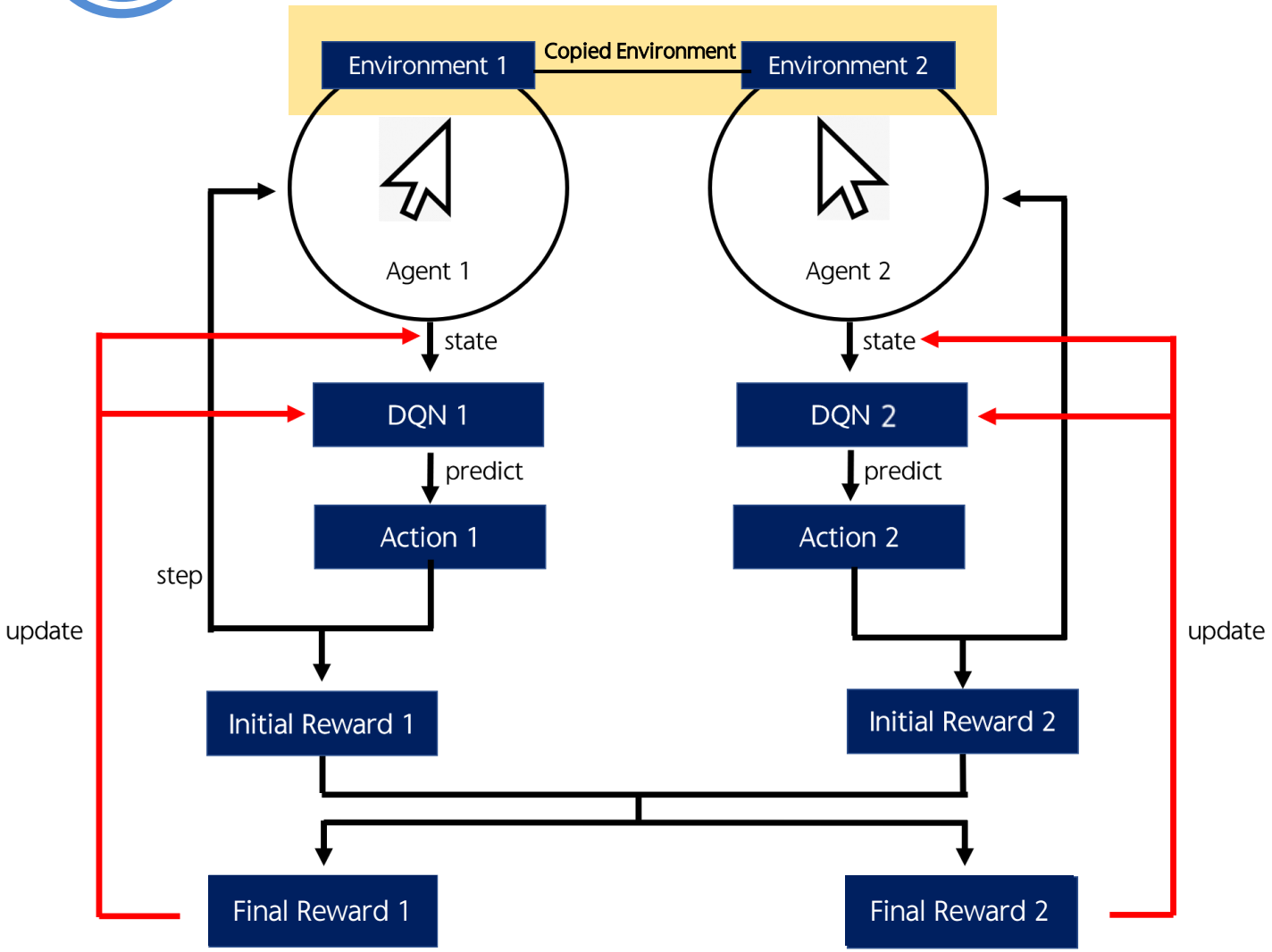
12개의 Human Factor

$P_{opponent}$ $V_{opponent}$

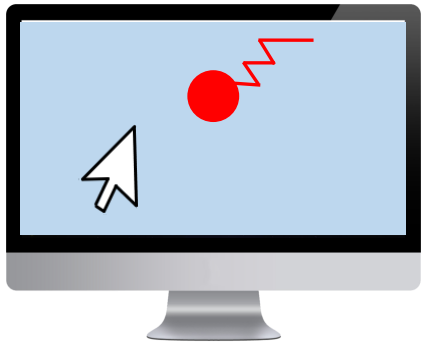
들을 고려하여 DQN에서 Reward를

최대화하는 action 선택

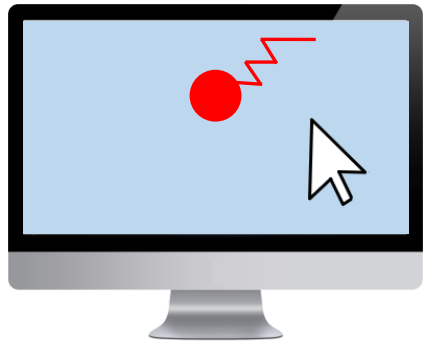
03 (1) 적대적 강화학습 환경 구축



① Environment



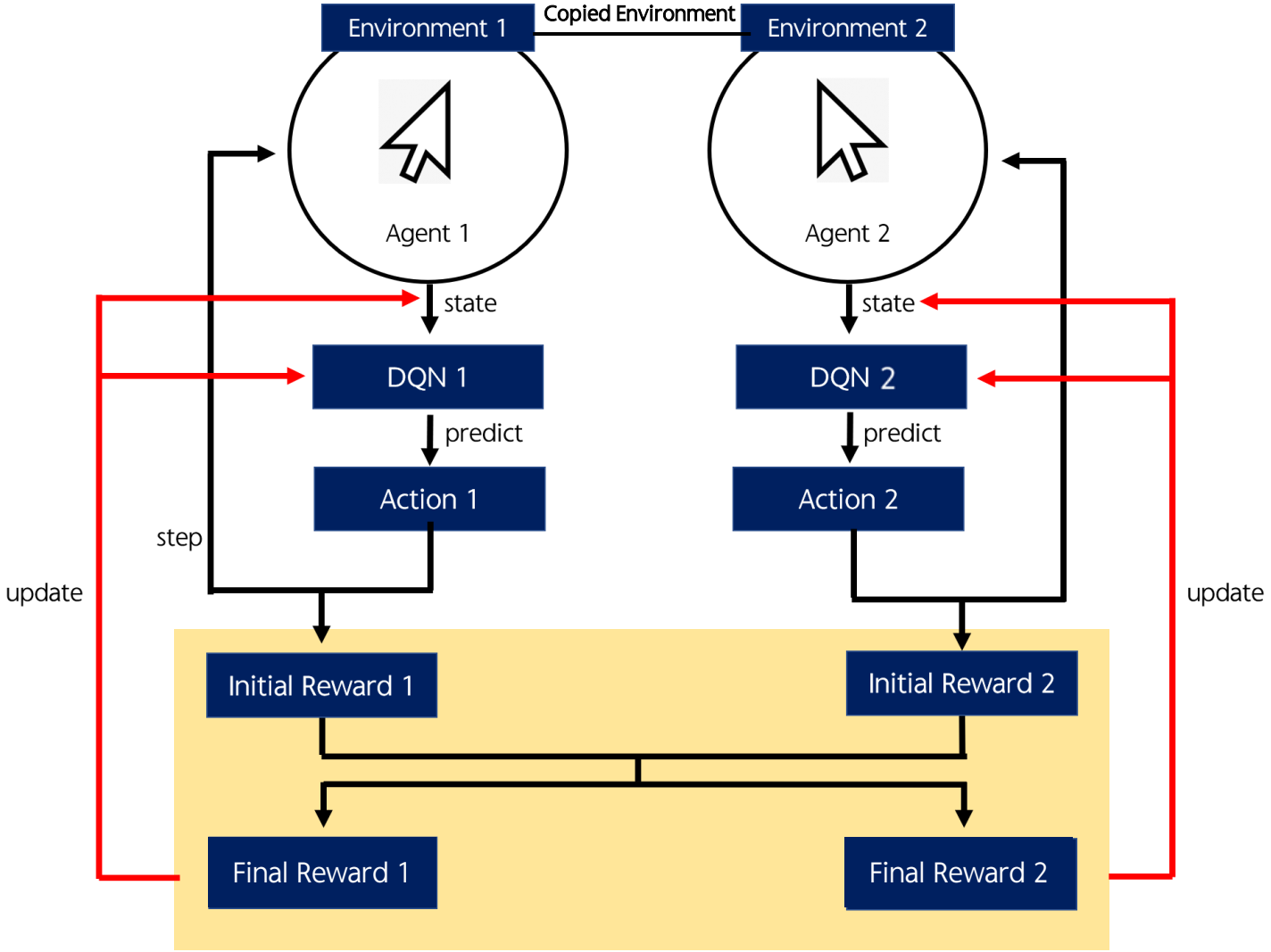
Environment 1



Environment 2

- 두 agent는 서로 같은 target을 생성하는 서로 다른 environment에서 학습
- 각 agent가 target을 취득(클릭)하는데 걸리는 시간, 클릭 성공여부 등을 측정

03 (1) 적대적 강화학습 환경 구축



② Reward

Agent 1

Initial Reward : R++

Final Reward : R+

Agent 2

Initial Reward : R++

Final Reward : R++

→ opponent가 target을 클릭하는 데 걸린 시간 및 클릭 성공 여부를 함께 고려해서 **다시 계산한 최종 Reward를 기준**으로 각 agent의 **policy를 업데이트**

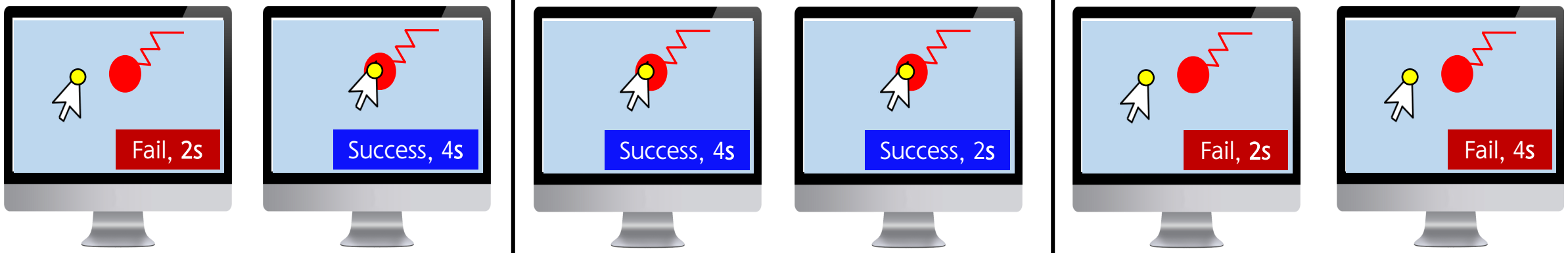
03 (1) 적대적 강화학습 환경 구축

② Reward = Click Reward – Motor Effort

Case	Click Reward	Motor Effort	Example
Case 1 (Click 시도 이전)	0	$-\sum_{t=t_0+T_p}^{t_0+T_p+T_h} \left\|\dot{\hat{\mathbf{v}}}_h[t]\right\|$	-
Case 2 (빠르게 Click 성공)	14	$-\sum_{t=t_0+T_p}^{t_0+T_p+T_h} \left\|\dot{\hat{\mathbf{v}}}_h[t]\right\|$	상대보다 먼저 클릭해서 성공한 경우, 상대가 먼저 실패한 후에 내가 성공한 경우
Case 3 (늦게 Click 성공)	9	$-\sum_{t=t_0+T_p}^{t_0+T_p+T_h} \left\|\dot{\hat{\mathbf{v}}}_h[t]\right\|$	상대가 먼저 성공한 이후에 내가 성공한 경우
Case 4 (Click Fail)	-1	$-\sum_{t=t_0+T_p}^{t_0+T_p+T_h} \left\|\dot{\hat{\mathbf{v}}}_h[t]\right\|$	시점에 관계없이 클릭에 실패한 경우

03 (1) 적대적 강화학습 환경 구축

③ Episode Termination = 두 Agent 가 모두 Click 기회를 소진하는 경우



Agent 1 Agent 2 | Agent 1 Agent 2 | Agent 1 Agent 2

Episode 종료 시점 : **4s** | Episode 종료 시점 : 4s | Episode 종료 시점 : **4s**

→ 두 Agent가 모두 기본적으로 Click Success를 목표로 하도록 설정

→ Motor Effort 만을 줄이는 방향으로 잘못 수렴되지 않도록 설정

04

연구 결과

(1) 동일한 Agent 간의 적대적 강화학습 결과

(2) 서로 다른 Agent 간의 적대적 강화학습 결과

Human Factor

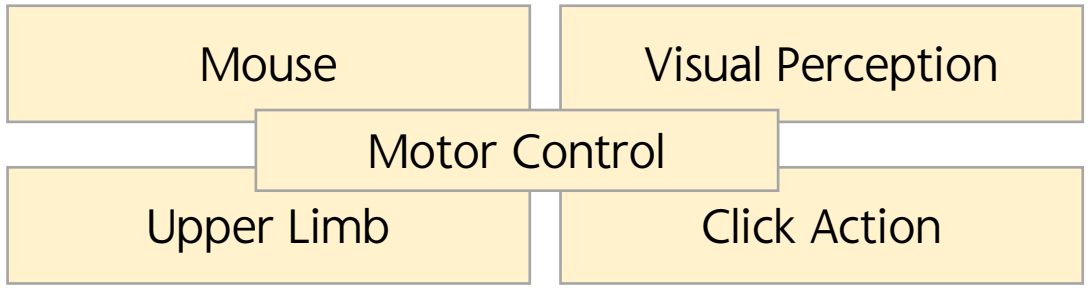
04 (1) 동일한 Agent 간의 적대적 강화학습 결과

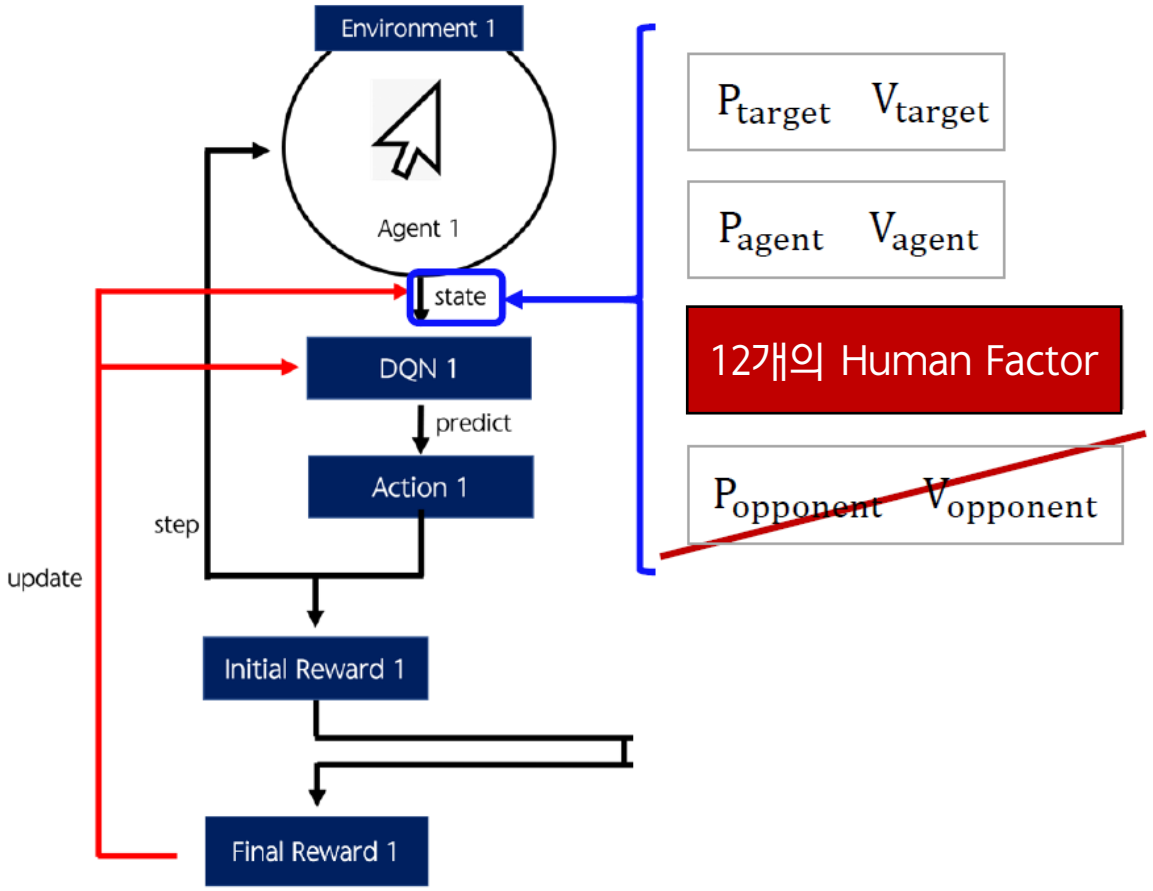
- Human Factor

→ 인간의 특성을 state에 반영하기 위한 변수

→ 5개의 Submodule, 12개의 Factor들로 구성

Variable	Description	Value	Ref	Module
T_p	Planning time interval	0.1 s	[11]	Motor control
n_v	Motor noise constant (parallel)	0.2	[44]	Upper limb
n_p	Motor noise constant (perpendicular)	0.02	[44]	Upper limb
l_{se}	Shoulder-to-elbow length	25.7 cm	[40]	Upper limb
l_{ew}	Elbow-to-wrist length	25.7 cm	[54]	Upper limb
l_{wh}	Wrist-to-hand length	6.43 cm	[40]	Upper limb
σ_v	Width of likelihood of visual speed perception	0.15	[60]	Visual perception
$f_{gain}()$	Mouse acceleration function	OS X 10.12	[14]	Mouse
c_σ	Precision of internal clock	0.09015	[41, 53]	Click action
c_μ	Implicit aim point	0.185	[41, 53]	Click action
ν	Drift rate	19.931	[41, 53]	Click action
δ	Visual encoding precision limit	0.399	[41, 53]	Click action





Do et al. 2021. A Simulation Model of Intermittently Controlled Point-and-Click Behaviour

03 (1) 동일한 Agent 간의 적대적 강화학습 결과

실험 1. 서로 동일한 Human Factor 값을 가지는 Agent 사이의 경쟁적 강화학습

Variable	Description	Value	Ref	Module
T_p	Planning time interval	0.1 s	[11]	Motor control
n_v	Motor noise constant (parallel)	0.2	[44]	Upper limb
n_p	Motor noise constant (perpendicular)	0.02	[44]	Upper limb
l_{se}	Shoulder-to-elbow length	25.7 cm	[40]	Upper limb
l_{ew}	Elbow-to-wrist length	25.7 cm	[54]	Upper limb
l_{wh}	Wrist-to-hand length	6.43 cm	[40]	Upper limb
σ_v	Width of likelihood of visual speed perception	0.15	[60]	Visual perception
$f_{gain}()$	Mouse acceleration function	OS X 10.12	[14]	Mouse
c_σ	Precision of internal clock	0.09015	[41, 53]	Click action
c_μ	Implicit aim point	0.185	[41, 53]	Click action
ν	Drift rate	19.931	[41, 53]	Click action
δ	Visual encoding precision limit	0.399	[41, 53]	Click action

고정

	기존 연구 (Do et al.)	Ours
Human Factor	동일한 값 사용	
학습방식	강화학습	적대적 강화학습

통제변인

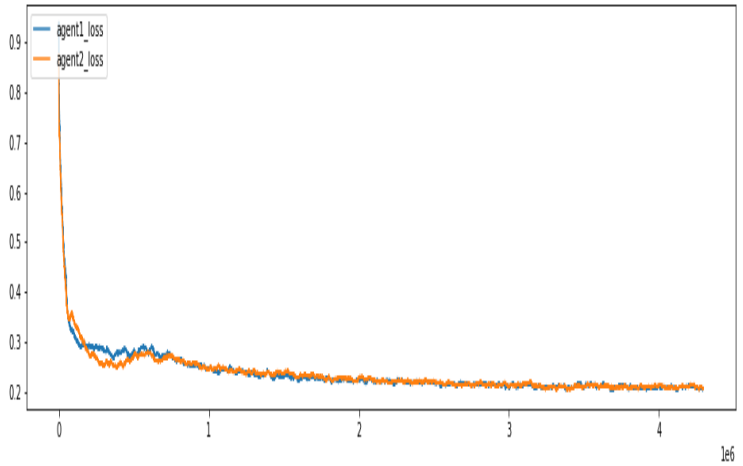
→ 적대적 강화학습을 사용하지 않았던 기존 연구와 **동일한 Human Factor** 값을 사용

조작변인 / 종속변인

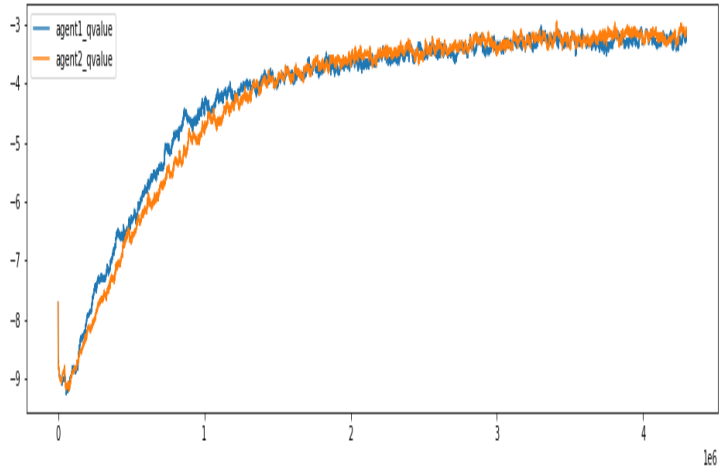
→ **적대적 강화학습** 이 **Policy** 변화에 주는 영향을 검증하는 실험

04 (1) 동일한 Agent 간의 적대적 강화학습 결과

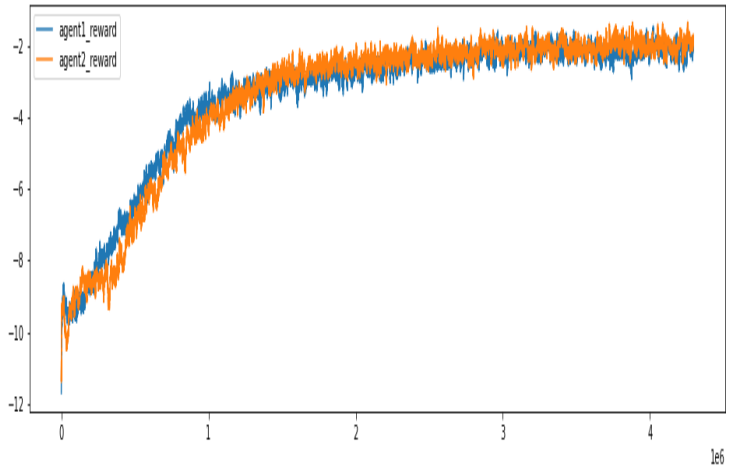
Loss



Q-Value



Reward

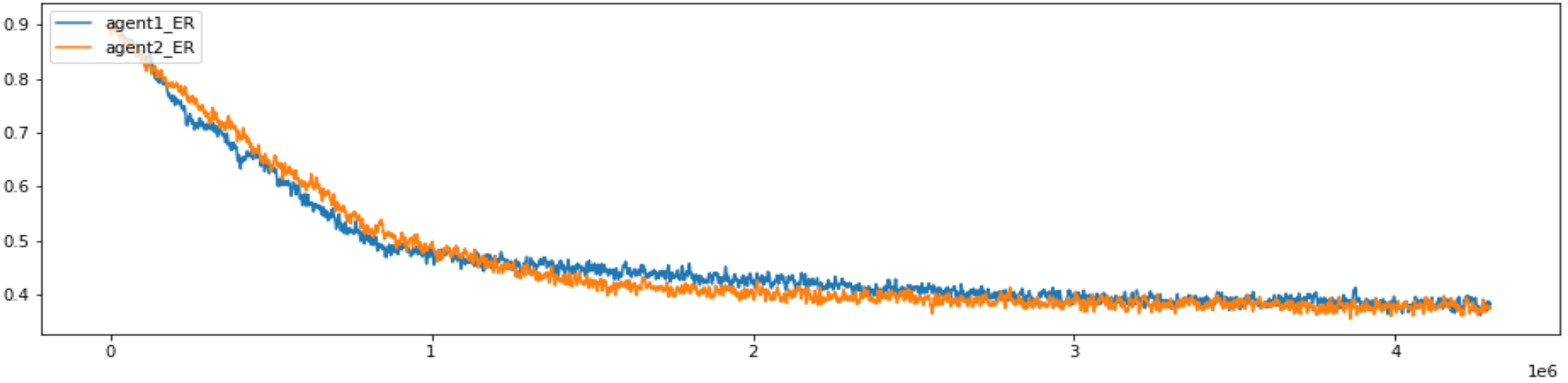


- 2주간 약 4.3M개 Episode 학습
- Loss는 단조 감소, Q-Value와 Reward는 단조 증가하는 바람직한 양상을 보임

Agent 1
Agent 2

04 (1) 동일한 Agent 간의 적대적 강화학습 결과

Click Failure Rate



- 두 Agent 모두 0.4로 수렴

Agent 1
Agent 2

04 (1) 동일한 Agent 간의 적대적 강화학습 결과

- Model Comparison & Evaluation

Adversarial Model (Ours) VS Non-Adversarial Model (Do et al, 2021)

Agent 1 == Agent 2 이므로 하나의 Agent 선택

→ 적대적 강화학습이 Policy 변화에 주는 영향 평가

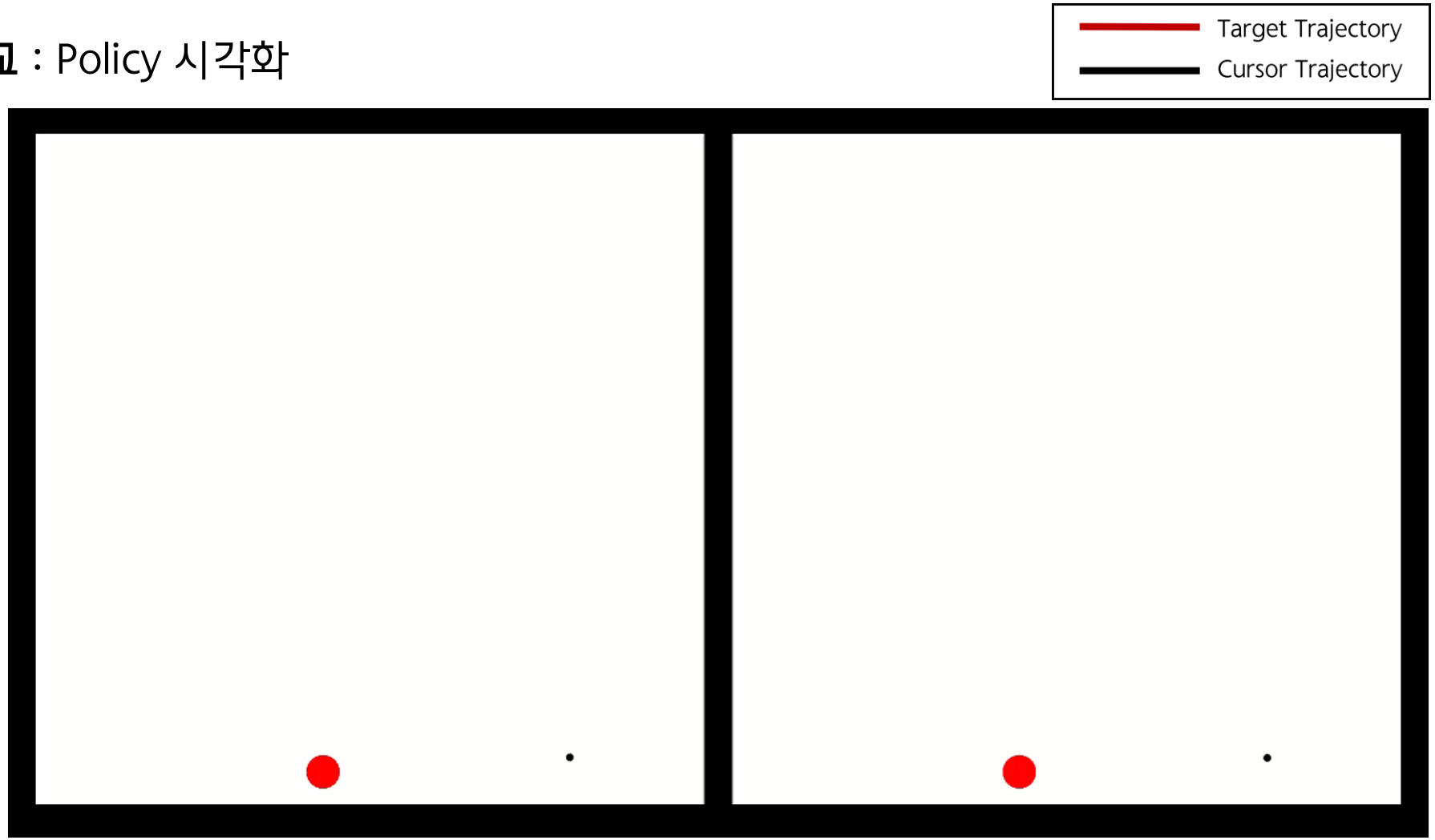
① **정성적 비교** : Policy 시각화

② **정량적 비교** : Trial Completion Time, Click Failure Rate 비교

(Agent가 클릭을 "시도"하는데 까지 걸리는 시간) (클릭 실패율)

04 (1) 동일한 Agent 간의 적대적 강화학습 결과

① **정성적 비교** : Policy 시각화

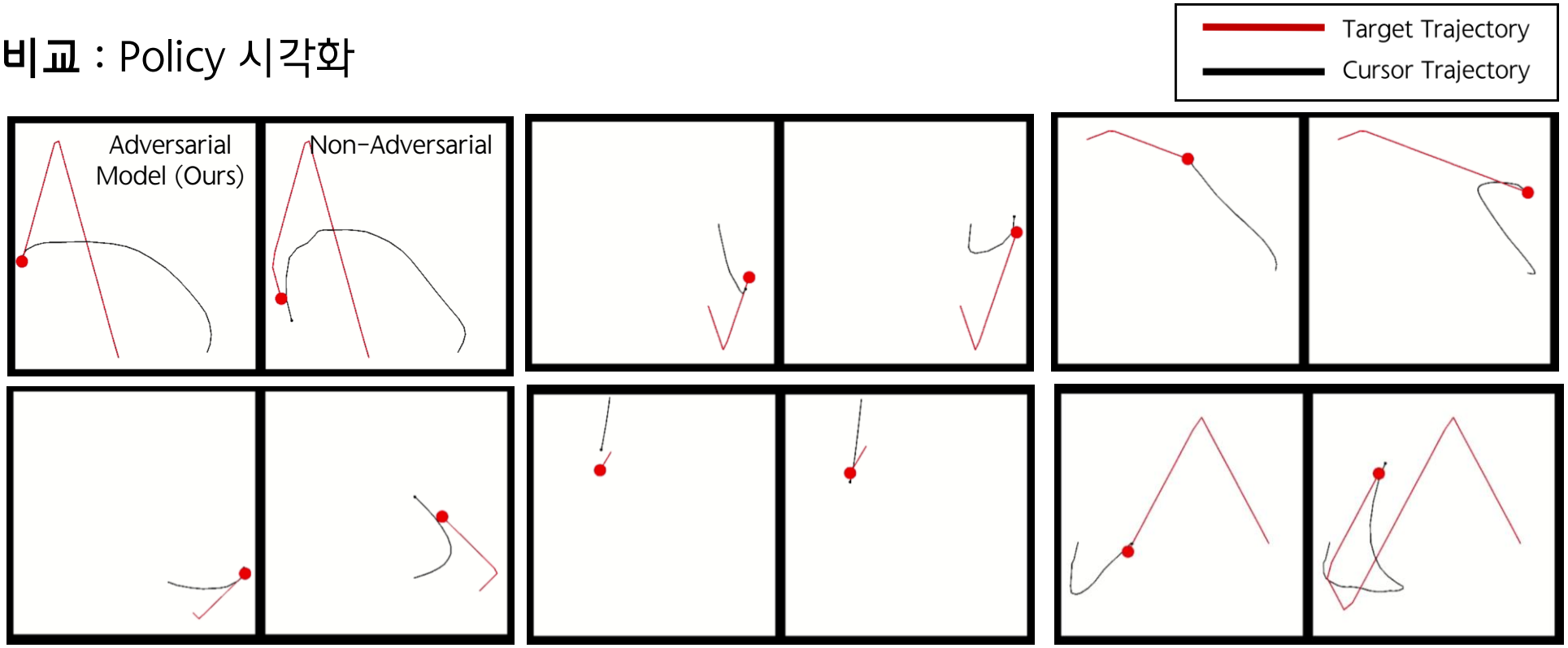


Adversarial Model (Ours)

Non-Adversarial Model (Do et al.)

04 (1) 동일한 Agent 간의 적대적 강화학습 결과

① **정성적 비교** : Policy 시각화



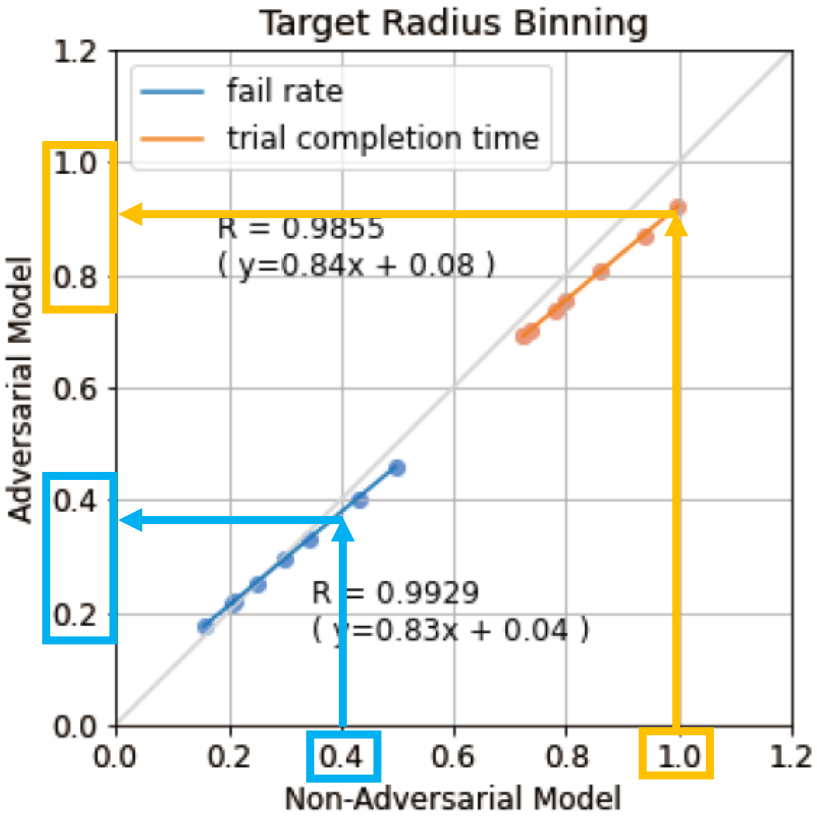
→ Target과 멀리 있을 때 두 모델의 Policy는 비슷한 양상

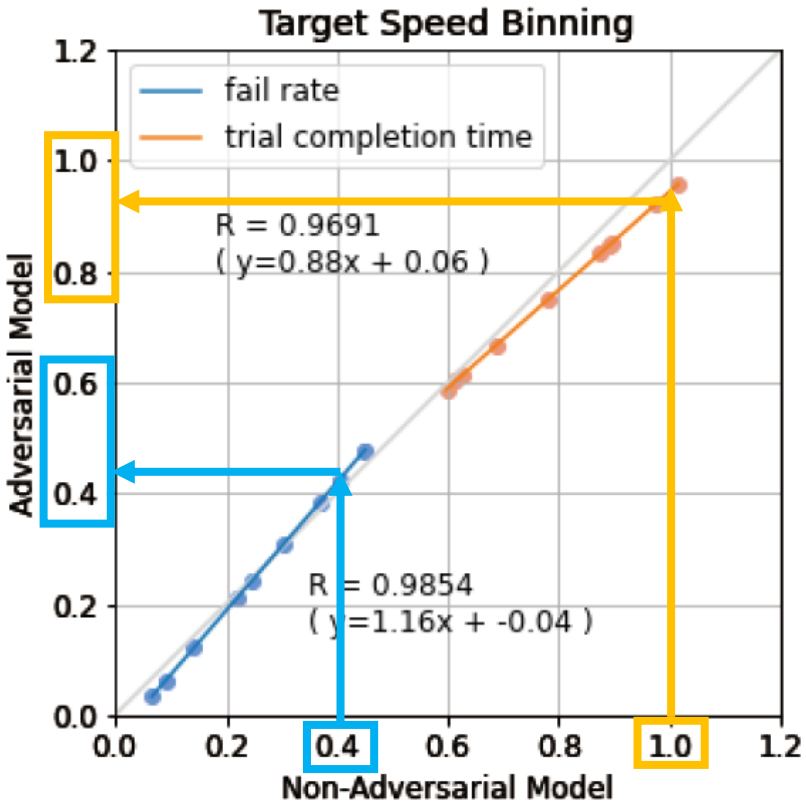
→ 적대적 모델은 **Target 근처로 접근 시** 짧은 Prediction-Horizon을 기준으로 **빨리 클릭을 시도하는 Policy**를 보임.

→ 반면, **기존 모델**은 Target의 이동 경로를 더 먼 미래까지 고려해서 클릭을 시도하는 Policy를 보임

04 (1) 동일한 Agent 간의 적대적 강화학습 결과

② 정량적 분석

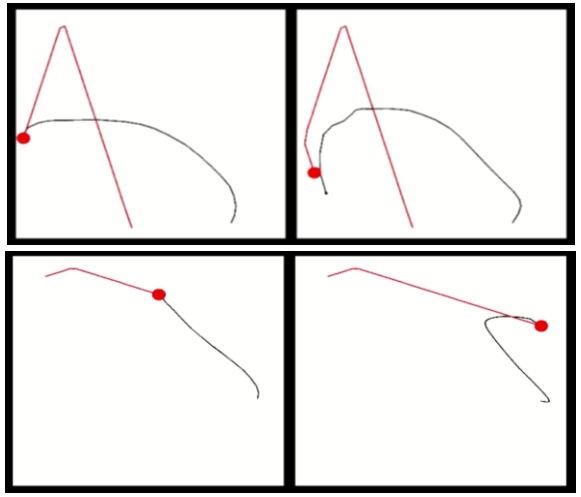


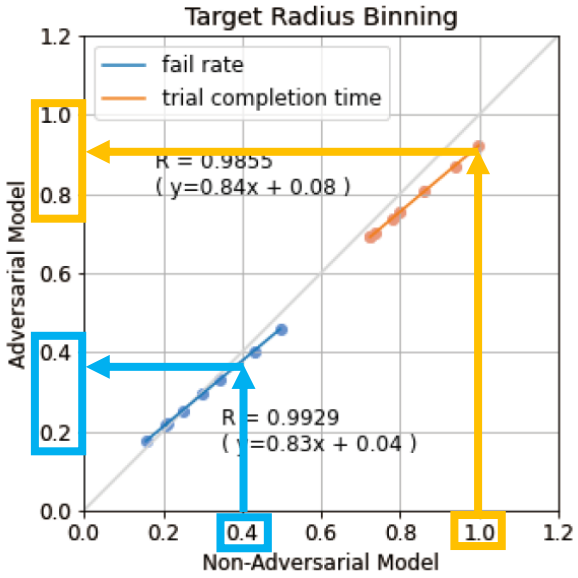


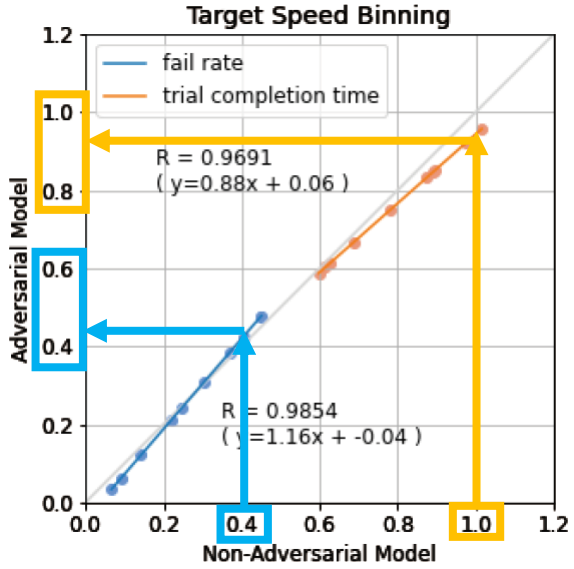
→ **Adversarial Model**이 Non-Adversarial Model에 비해 **Trial Completion Time이 짧은 경향**을 보임

04 (1) 동일한 Agent 간의 적대적 강화학습 결과

③ 결론







→ 두 모델의 **초기 Policy**는 비슷

→ Target과 가까워졌을 때 **Adversarial Model**은 짧은 Prediction Horizon을 가지고 빨리 클릭해버리는 Policy

→ **Adversarial Model**의 Non-Adversarial Model에 비해 **Trial Completion Time이 짧음**

→ 반면, Non-Adversarial Model은 Target의 이동 경로를 더 먼 미래까지 고려해서 클릭을 시도하는 Policy를 보임

04

연구 결과

(1) 동일한 Agent 간의 적대적 강화학습 결과

(2) 서로 다른 Agent 간의 적대적 강화학습 결과

03 (2) 적대적 강화학습 환경 구축 (개별)

실험 2. 서로 다른 Human Factor 값을 가지는 Agent 사이의 경쟁적 강화학습

Agent 1

(Improved **Decision-Making Skill** Agent)

변수		Value	Module
c_μ	Precision of internal clock	0.185 → 0.3	Click action
c_σ	Implicit aim point	0.09015 → 0.06	Click action
ν	Drift rate	19.931 → 40	Click action
δ	Visual encoding precision limit	0.399 → 0.25	Click action

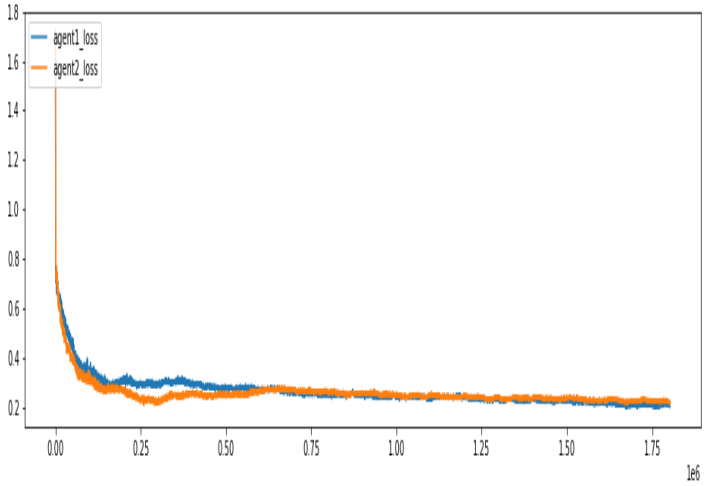
Agent 2

(Improved **Motor Execution** Agent)

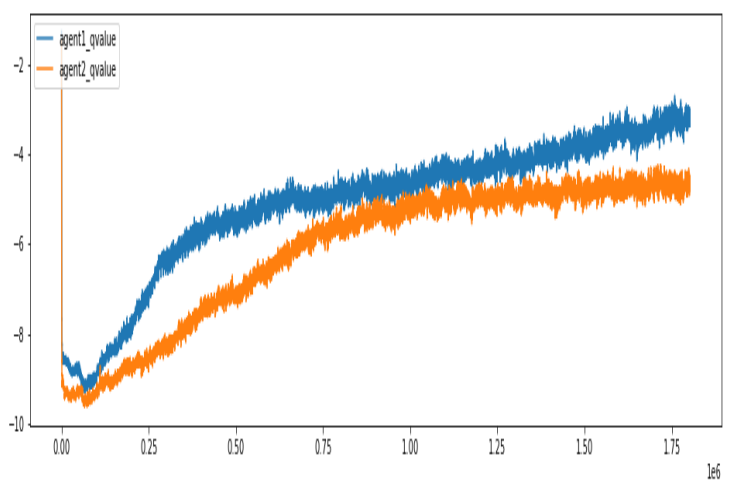
변수		Value	Module
n_v	Motor noise constant (parallel)	0.2 → 0.24	Upper limb
n_p	Motor noise constant (perpendicular)	0.02 → 0.024	Upper limb
σ_v	Width of likelihood of visual speed perception	0.15 → 0.18	Visual Perception

04 (2) 서로 다른 Agent 간의 적대적 강화학습 결과

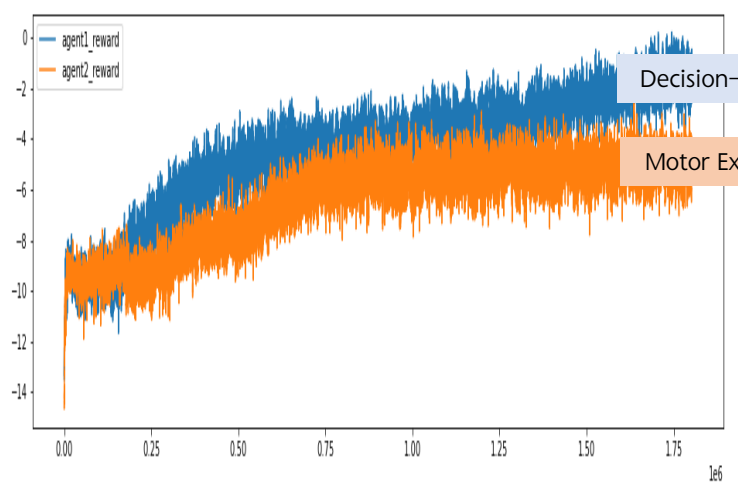
Loss



Q-Value



Reward



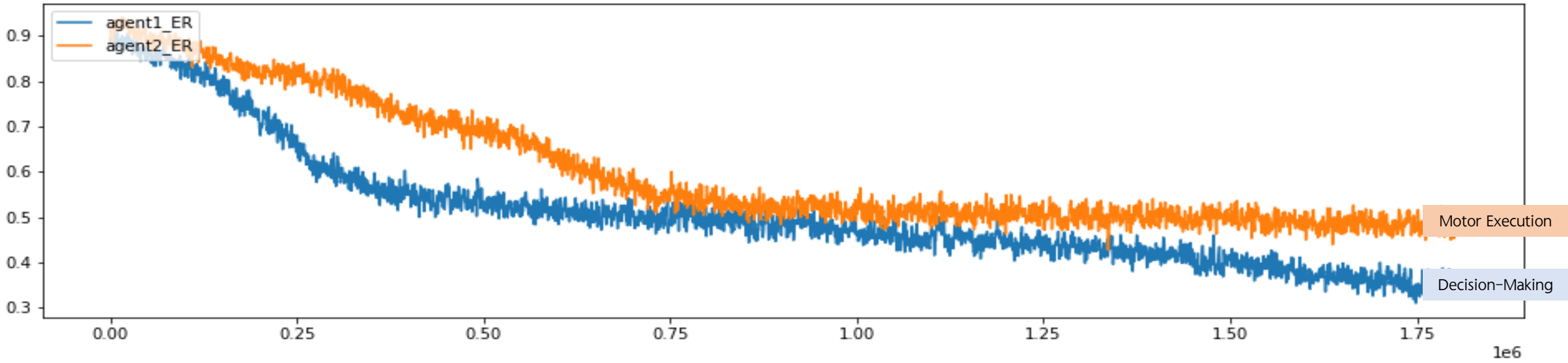
Decision-Making

Motor Execution

- 1주간 약 1.8M개 Episode 학습
- Loss는 단조 감소, Q-Value와 Reward는 단조 증가하는 바람직한 양상을 보임
- Improved Decision-Making Skill Agent가 Reward를 더 빠르게 누적하는 양상을 보임

04 (2) 서로 다른 Agent 간의 적대적 강화학습 결과

Click Failure Rate



- Improved Decision-Making Skill Agent 가 더 낮은 값(0.35)으로 수렴

04 (2) 서로 다른 Agent 간의 적대적 강화학습 결과

- Model Comparison & Evaluation

Improved **Decision-Making Skill** Agent VS Improved **Motor Execution** Agent

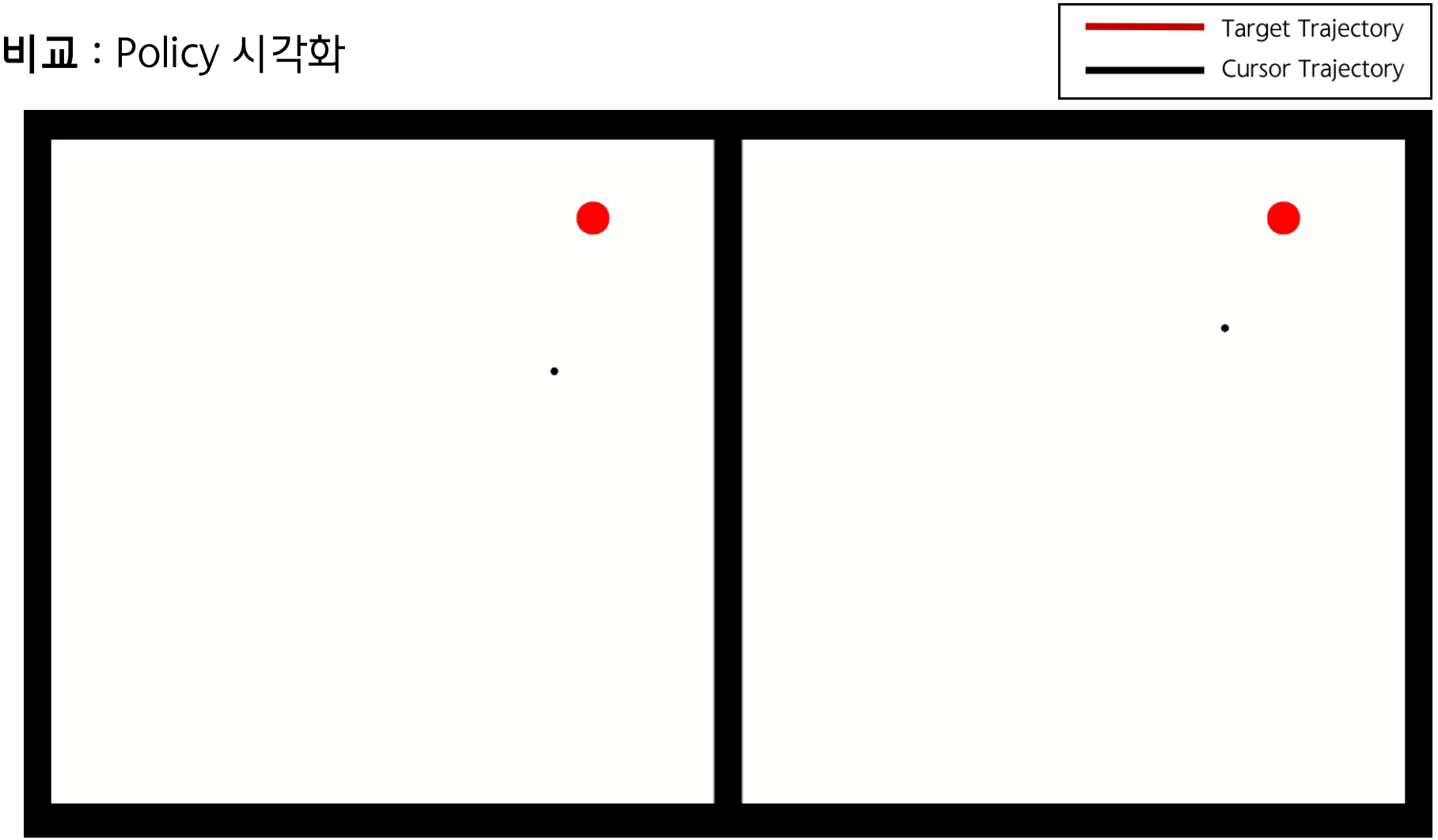
→ Point-and-Click Task를 수행하는 데 **영향력이 더 큰 Human Factor를 확인**

① **정성적 비교** : Policy 시각화

② **정량적 비교 :** Trial Completion Time, Click Failure Rate 비교

04 (2) 서로 다른 Agent 간의 적대적 강화학습 결과

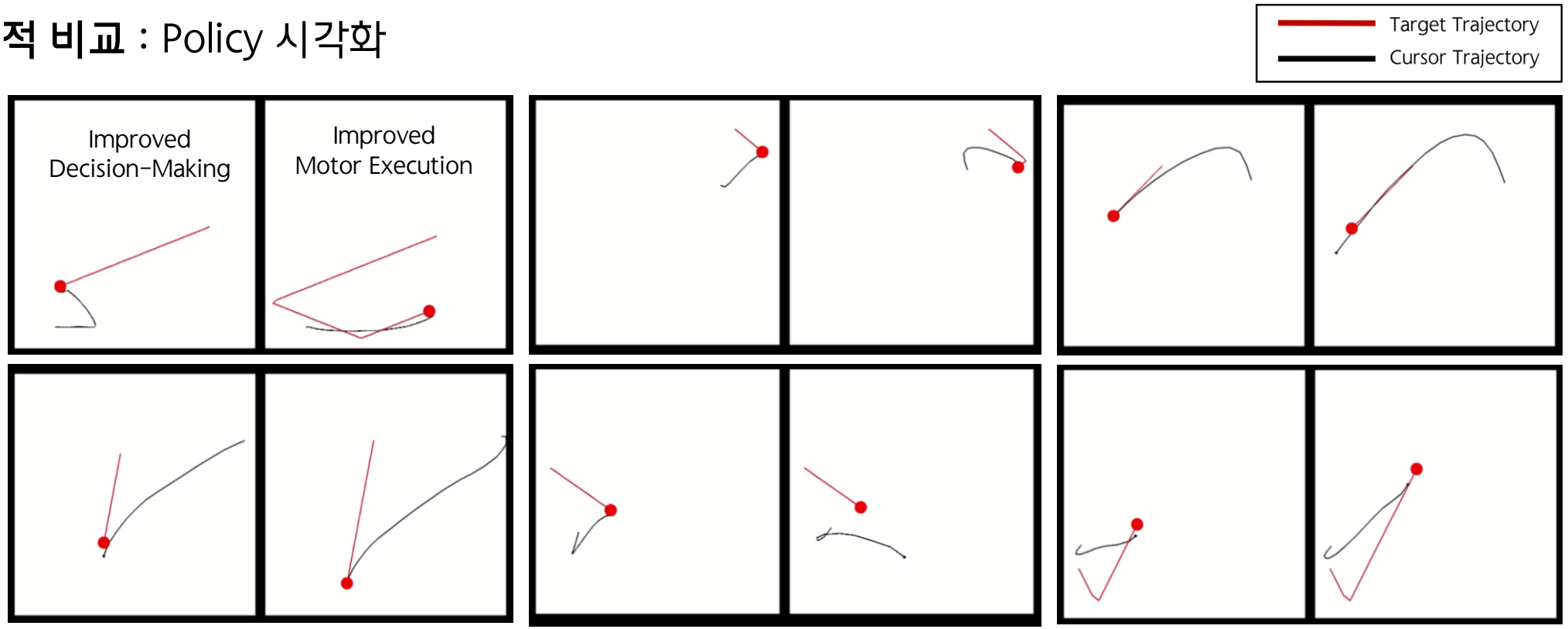
① **정성적 비교** : Policy 시각화



Improved Decision-Making Skill Agent　　Improved Motor Execution Agent

04 (2) 서로 다른 Agent 간의 적대적 강화학습 결과

① **정성적 비교** : Policy 시각화

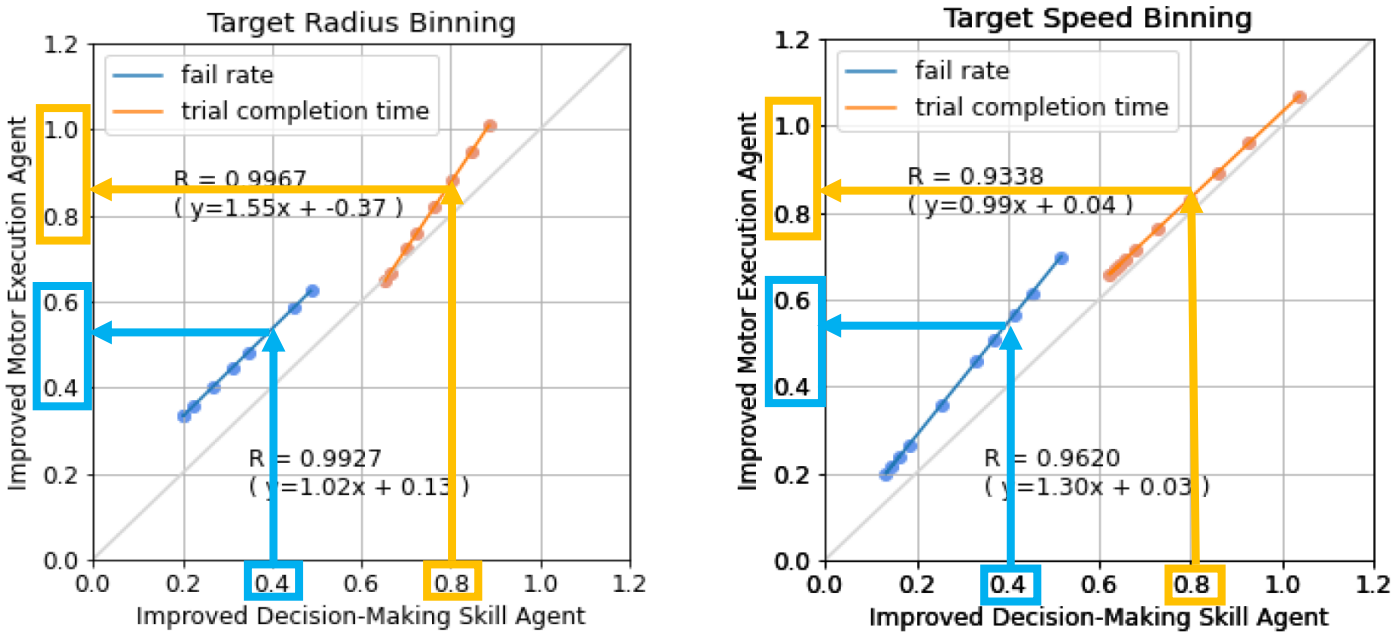


→ Improved **Decision-Making Skill** Agent : Target의 **현재 위치로 직진**

→ Improved **Motor Execution** Agent : **먼 미래**까지 (**boundary에서 튕기는 것까지**) 고려해서 Target에 대한 **상대속도를 0**으로 맞추려고 **곡선으로 접근하는 경향**

04 (2) 서로 다른 Agent 간의 적대적 강화학습 결과

② 정량적 분석 및 결론



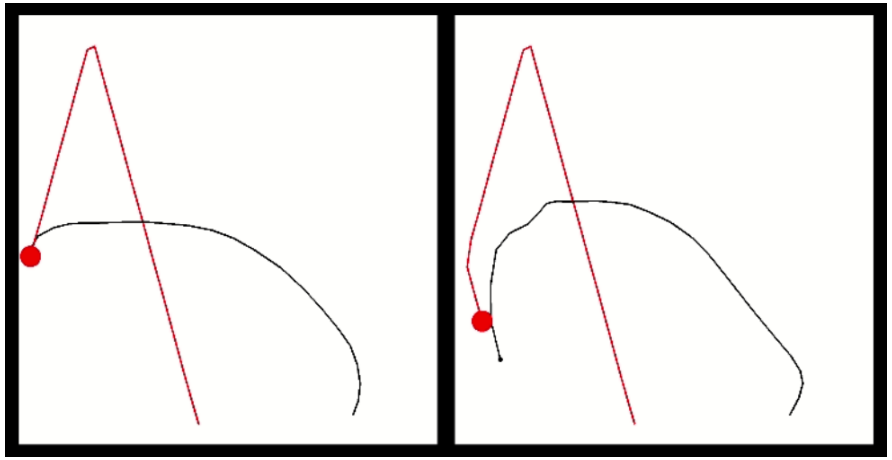
→ Improved Decision-Making Skill Agent가 Trial Completion Time과 Click Failure Rate가 모두 작음

= 더 빠르고 정확하게 클릭

= Point-and-Click Task 수행 시 Decision-Making Skill이 큰 영향을 줌

05 결론

Conclusion & Contribution

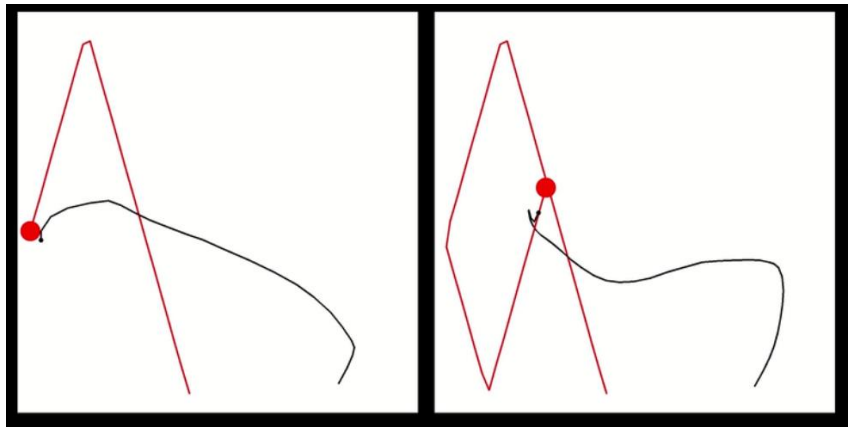


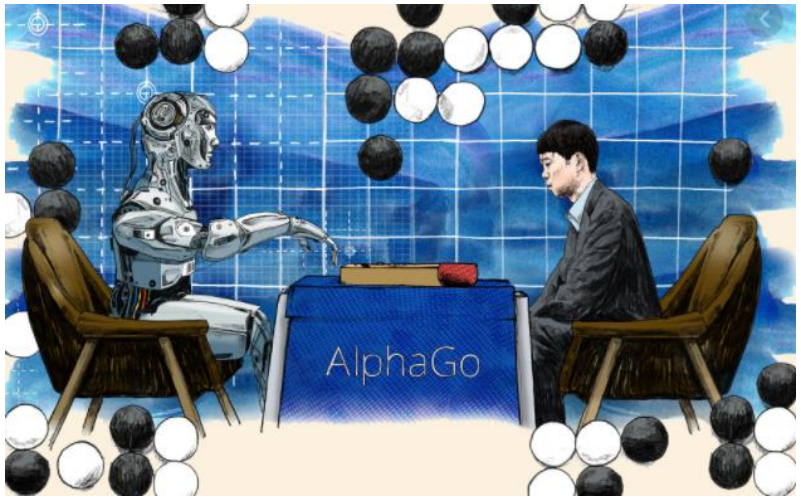
(1) 적대적 모델의 Optimal Policy는 기존 모델의 Optimal Policy와 정성적, 정량적인 차이 존재

- Trial Completion Time을 낮추는 방향으로 최적화
- 실제 적대적 환경에서의 Optimal Policy / 전략을 시사

05 결론

Conclusion & Contribution





(2) Point-and-Click Task 수행 시 **Decision-Making Skill의 중요성을 강화학습으로 증명**

- E. Park, B. Lee. 2020. An Intermittent Click Planning Model

- Gamer Group VS Non-Gamer Group : Decision-Making Skill Related Free Parameter 차이 존재

(3) **프로게이머 등을 위한 가이드라인, 인간과 겨루는 Point-and-Click Agent 개발** 등에 폭넓게 활용