

Reinforcement Learning

# 나는 강화학습으로 축구한다

Google Research와 Manchester City F.C.의  
인공지능 축구 프로젝트를 대한민국에서  
재조명하고 직접 구현해보는 캠프



SESSION

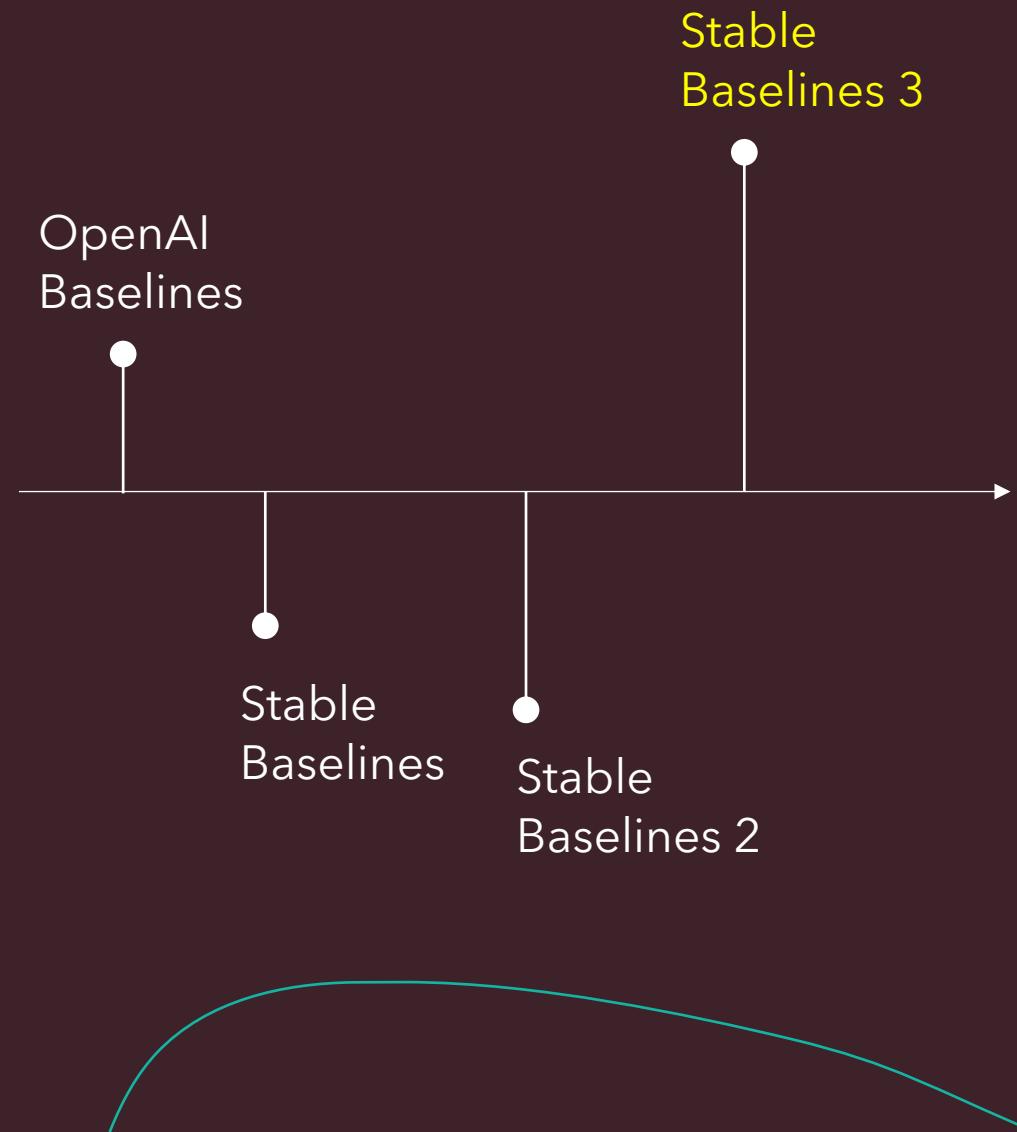
17

# Copyright Notice

These slides are available for educational purposes. You may not use or distribute these slides for commercial purposes. You may make copies of these slides and use or distribute them for educational purposes as long as you cite the author as the source of the slides.

# Stable Baselines 3 (SB3)

Features	Stable-Baselines	OpenAI Baselines
State of the art RL methods	✓ (1)	✓
Documentation	✓	✗
Custom environments	✓	✓
Custom policies	✓	— (2)
Common interface	✓	— (3)
Tensorboard support	✓	— (4)
Ipython / Notebook friendly	✓	✗
PEP8 code style	✓	✓ (5)
Custom callback	✓	— (6)



# Proximal Policy Optimization (PPO)

```
import gfootball.env as football_env
from stable_baselines3 import PPO

env = football_env.create_environment(
    env_name="academy_empty_goal_close",
    representation="simple115v2"
)
env.reset()

model = PPO("MlpPolicy", env, verbose=1)

model.learn(total_timesteps=10_000)

model.save("model-ppo.zip")

env.close()
```

$$\text{Maximize } J(\boldsymbol{\theta}) = \mathbb{E}_{\tau \sim \pi_{\boldsymbol{\theta}}} [G_0]$$

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \mathbb{E}_{\tau \sim \pi_{\boldsymbol{\theta}}} \left[ \sum_{t=0}^{T-1} \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(\mathbf{a}_t | \mathbf{s}_t) G_t \right] = \frac{1}{N} \sum_{i=1}^N \sum_{t=0}^{T-1} \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(\mathbf{a}_t^i | \mathbf{s}_t^i) G_t^i$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \alpha \frac{1}{N} \sum_{i=1}^N \sum_{t=0}^{T-1} \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(\mathbf{a}_t^i | \mathbf{s}_t^i) G_t^i$$

$$\mathbf{a} \sim \pi_{\boldsymbol{\theta}}(\cdot | \mathbf{s}_t)$$

# DQN, A2C, PPO

```
import gfootball.env as football_env
from stable_baselines3 import DQN

env = football_env.create_environment(...)

env.reset()

model = DQN("MlpPolicy", env, verbose=1)

model.learn(...)

model.save("...")

env.close()
```

```
import gfootball.env as football_env
from stable_baselines3 import A2C

env = football_env.create_environment(...)

env.reset()

model = A2C("MlpPolicy", env, verbose=1)

model.learn(...)

model.save("...")

env.close()
```

```
import gfootball.env as football_env
from stable_baselines3 import PPO

env = football_env.create_environment(...)

env.reset()

model = PPO("MlpPolicy", env, verbose=1)

model.learn(...)

model.save("...")

env.close()
```

# Hands-On Exercise

PPO로 학습시켜보기

LAB  
06



Live

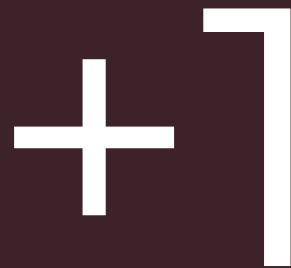


# 당근과 채찍질



# Reward Design

Decides what the agent ultimately optimizes



득점



실점

Environment 만들 때 정한  
궁극적으로 Agent가 성취하려는 목적

# Sparse Rewards

골이 너무 적게 나온다...

The NBA mobile application interface displays the latest game scores and highlights. The top navigation bar shows the NBA logo. Below it, there are four tabs: GAMES, STANDINGS, STATS, and PLAYERS. The GAMES tab is currently selected, showing the following game results:

Team	Score	Status	Action
Knicks	97	Q4 - 00:23	
Mavericks	114		
Spurs	123	Final Today	
Jazz	110		
Cavaliers	104	Final Today	
Thunder	136		

Below the scores, there are two additional rows of game cards:

Team	Score	Status	Action
76ers	33	Q2 - 12:00	
Pacers	30		
Wizards	106	Final Today	
Clippers	110		
Nets		Today 9:30AM	
Suns			

The Premier League mobile application interface displays the latest match results and highlights. The top navigation bar shows the Premier League logo and the text "2025-26 Season". Below it, there are four tabs: MATCHES, NEWS, STANDINGS, and STATS. The MATCHES tab is currently selected, showing the following match results for December 31, 2025:

Date	Match	Week	Match	Week
Dec 31, 2025 · FT	Burnley vs Newcastle	1	Chelsea vs Bournemouth	2
Dec 31, 2025 · FT	West Ham vs Brighton	2	Nottingham vs Everton	0
Dec 31, 2025 · FT	Manchester United vs Wolverhampton	1	Arsenal vs Aston Villa	4

Below the match results, there are two additional rows of match cards:

Match	Week	Match	Week
Burnley vs Newcastle	3	Chelsea vs Bournemouth	2
West Ham vs Brighton	2	Nottingham vs Everton	2
Manchester United vs Wolverhampton	1	Arsenal vs Aston Villa	1

# Reward Design vs. Reward Shaping

- Scoring: +1, -1
  - Learning becomes very slow
  - 당근도 안 주고 채찍질도 안 하는...
- Reward Shaping
  - Instead of giving the agent a reward only at the end (e.g. scoring a goal), you provide additional intermediate rewards that encourage progress toward the goal.
  - Agent receives feedback earlier.
  - Less exploration burden.

힌트주기? 빵가루 떨어뜨려놓기?

# Examples of Reward Shaping

- Acquiring Possession
- Moving Ball Closer to Opponent Goal
- Losing Possession
- Staying Longer in the Opponent's Side
- ...

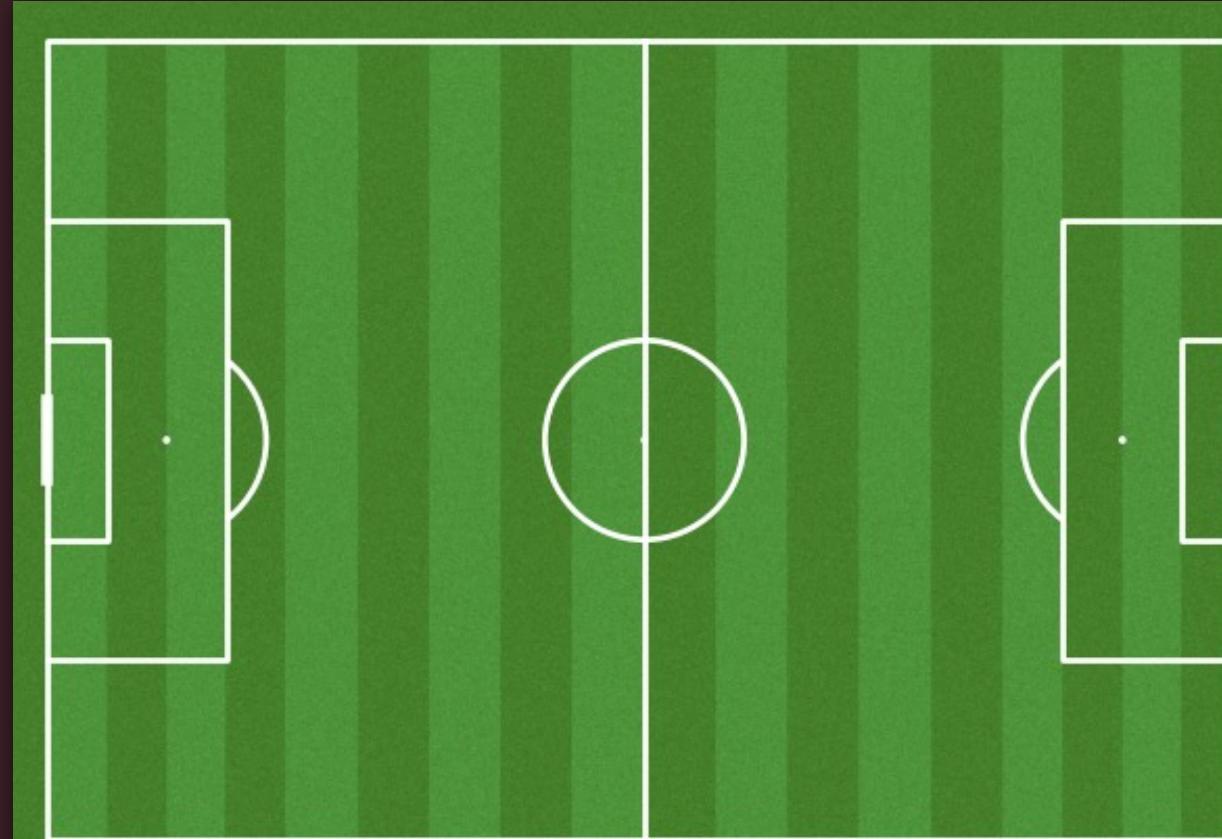
```
env = football_env.create_environment(  
    env_name="academy_run_to_score",  
    rewards="scoring"  
)
```

# Scoring



- ↗

실점



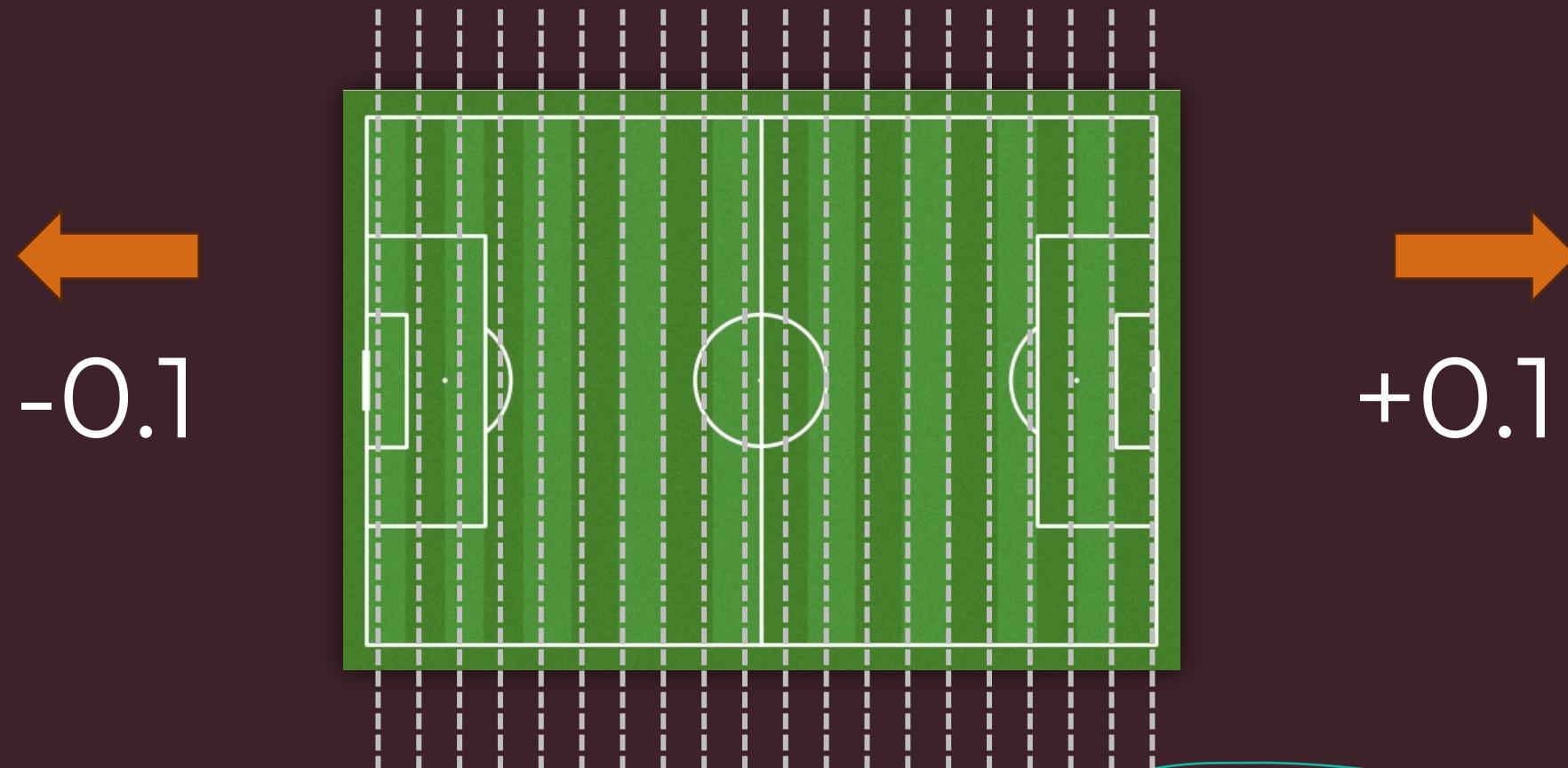
+ ↗

득점

```
env = football_env.create_environment(  
    env_name="academy_run_to_score",  
    rewards="scoring,checkpoints"  
)
```

# Checkpoints

(GRF Built-in Reward Shaping)



(the lines are arbitrarily drawn for the illustrative purpose)

# Other Issues

- Reward Hacking
  - Cheating
- Overfitting
  - Fails to generalize
- Conflicting Incentives
  - Dribble vs. Shot

# Applying SB3's PPO



academy\_run\_to\_score\_with\_keeper