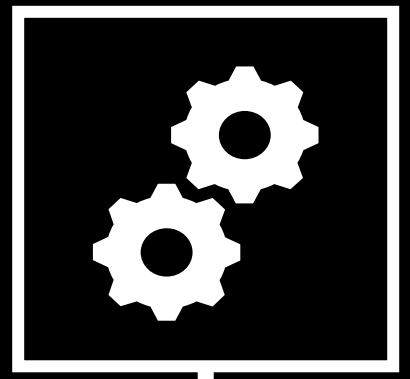
강화학습 개론
&
인공지능 툴 사용법

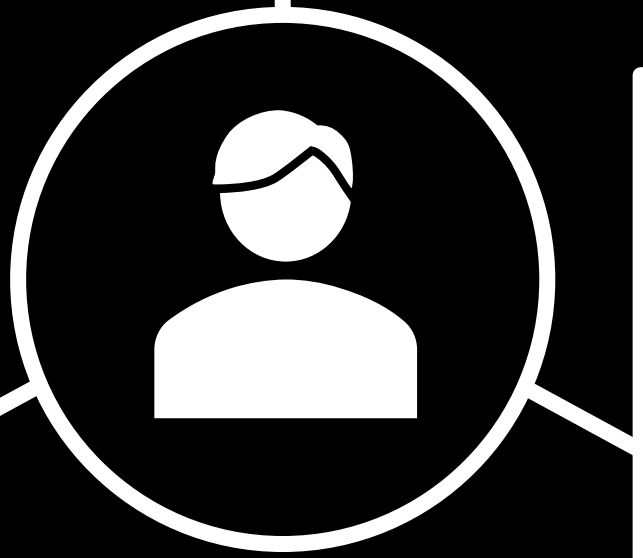


항공우주전투발전단
자원분석과 중위 이형진

소개

경기대 기계공학/학사





국방로봇 경진대회



Google AI 자격증

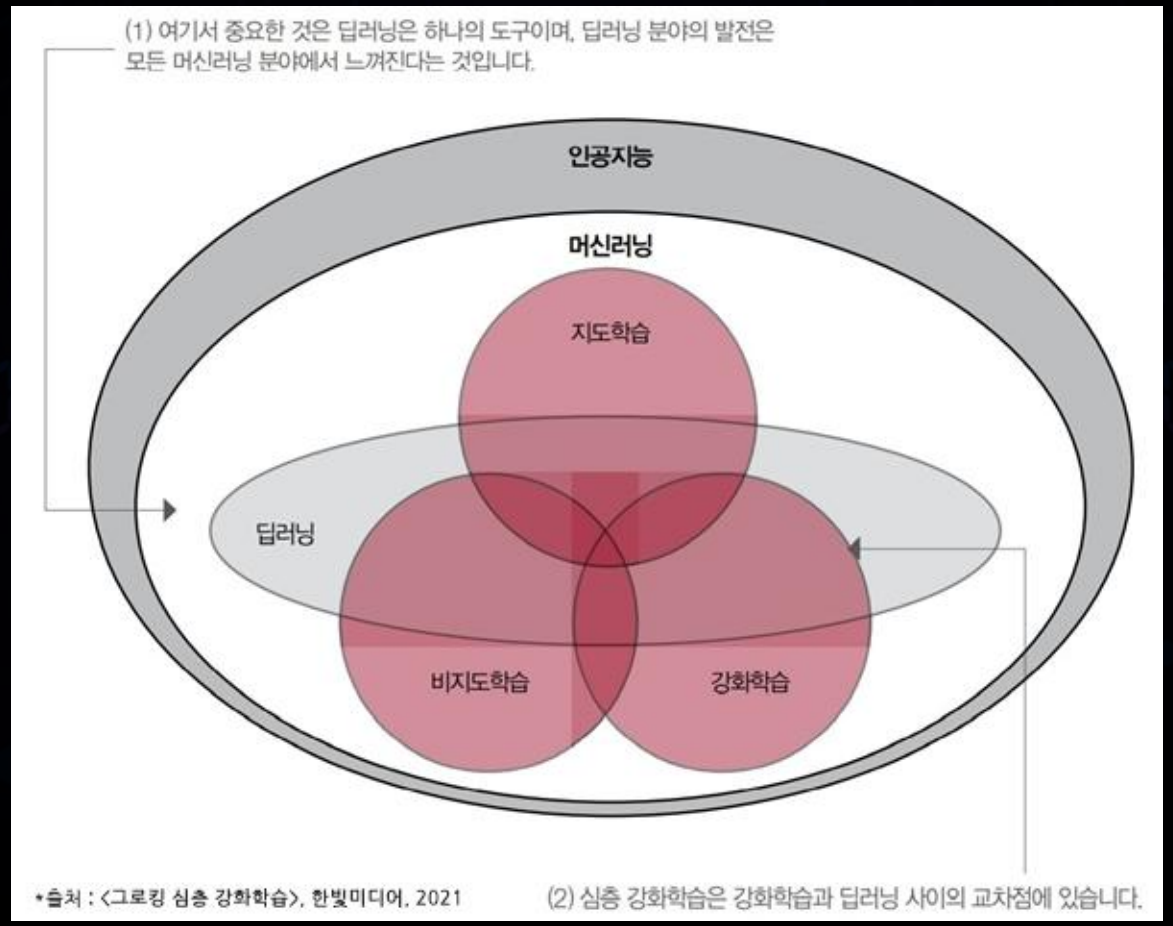


목차

- 강화학습 개론
  - 강화학습 용어
  - 하이퍼파라미터, 병렬학습
  - 사업단에서 고려할점

- 인공지능 툴 사용법(Cursor)
  - Cursor
  - Cursor + Obsidian

강화학습 개론


(1) 여기서 중요한 것은 딥러닝은 하나의 도구이며, 딥러닝 분야의 발전은 모든 머신러닝 분야에서 느껴진다는 것입니다.
인공지능
머신러닝
지도학습
딥러닝
비지도학습
강화학습
*출처 : <그로킹 심층 강화학습>, 한빛미디어, 2021
(2) 심층 강화학습은 강화학습과 딥러닝 사이의 교차점에 있습니다.

강화학습 개론





강화학습 개론



강화학습 개론

강화학습 수학

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

$$\text{PPO: } \max_\theta \hat{\mathbb{E}}_t \left[\min \left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\text{old}}(a_t|s_t)} \hat{A}_t, \text{clip} \left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\text{old}}(a_t|s_t)}, 1-\epsilon, 1+\epsilon \right) \hat{A}_t \right) \right]$$

⋮

강화학습 용어

환경

에이전트

상태

행동

정책

보상

가치함수

강화학습 용어

환경 → 에이전트 → 상태 → 행동 → 정책 → 보상 → 가치함수

환경

(1,4)	(2,4)	(3,4)	(4,4)
(1,3)	(2,3)	(3,3)	(4,3)
(1,2)	(2,2)	(3,2)	(4,2)
(1,1)	(2,1)	(3,1)	(4,1)

게임엔진

unity

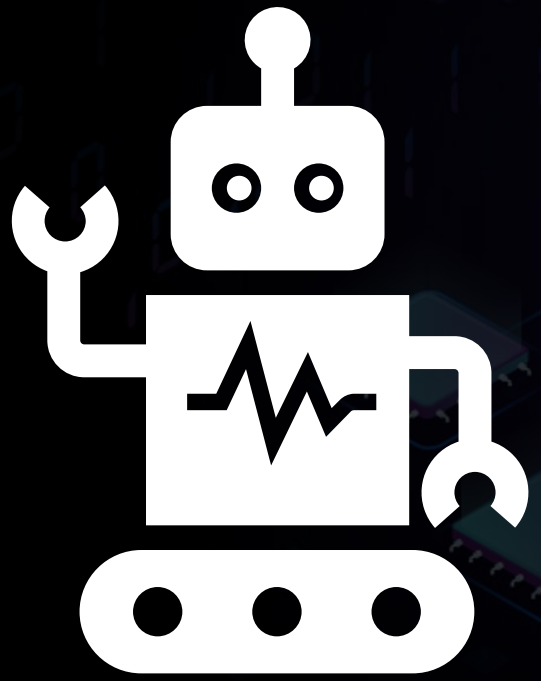
UNREAL ENGINE

GODOT Game engine

NVIDIA Omniverse

강화학습 용어

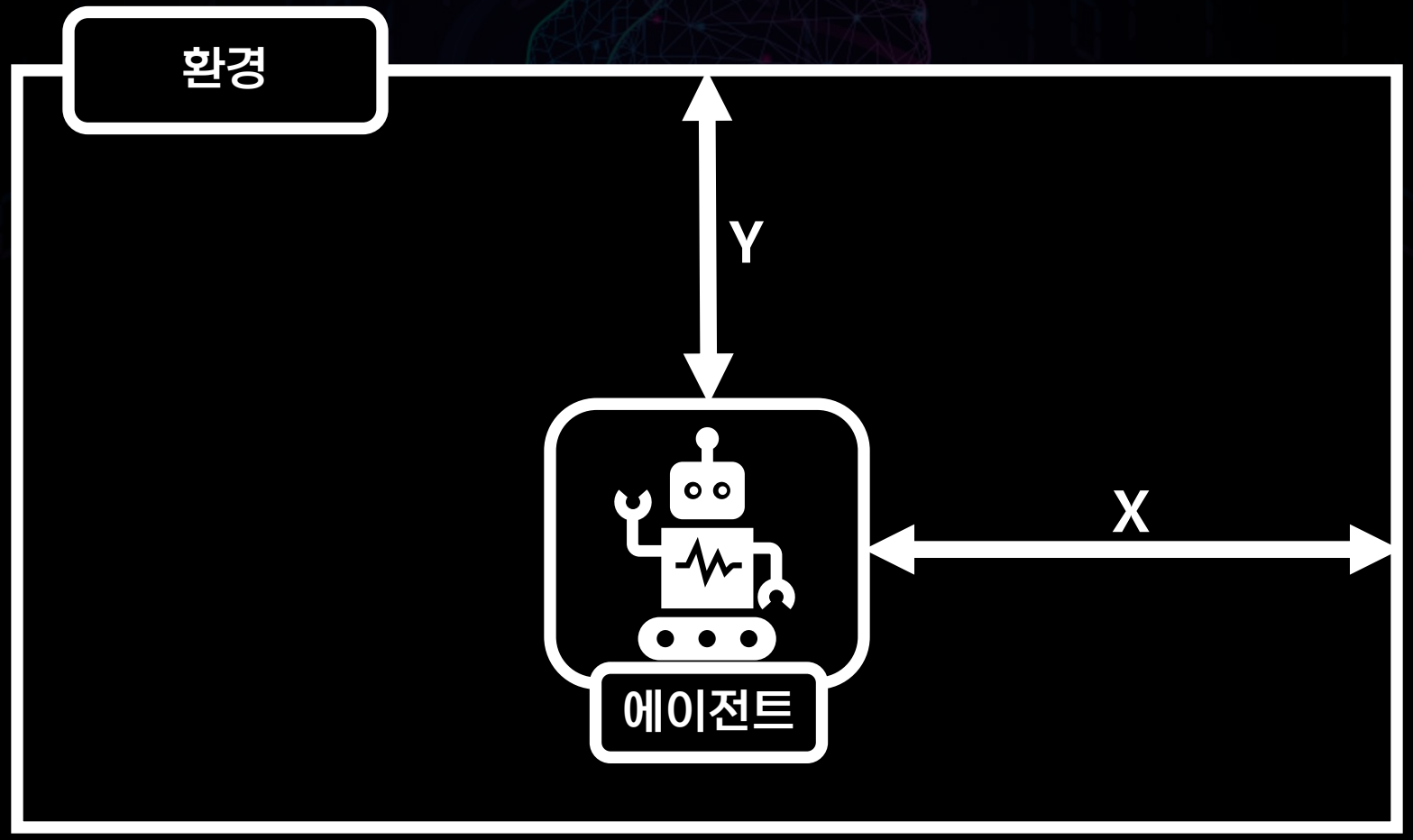
환경 → **에이전트** → 상태 → 행동 → 정책 → 보상 → 가치함수





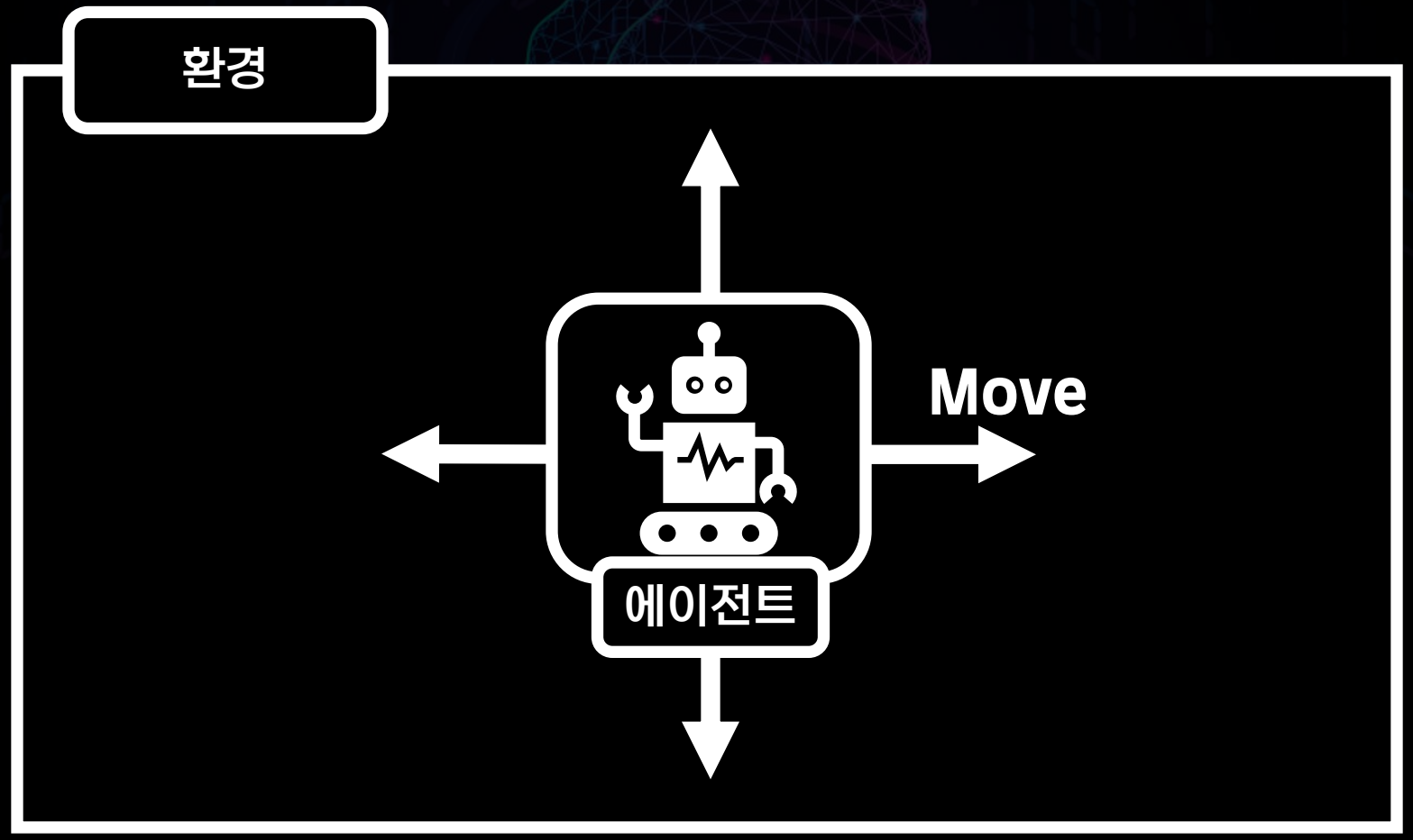
강화학습 용어

환경 → 에이전트 → 상태 → 행동 → 정책 → 보상 → 가치함수



강화학습 용어

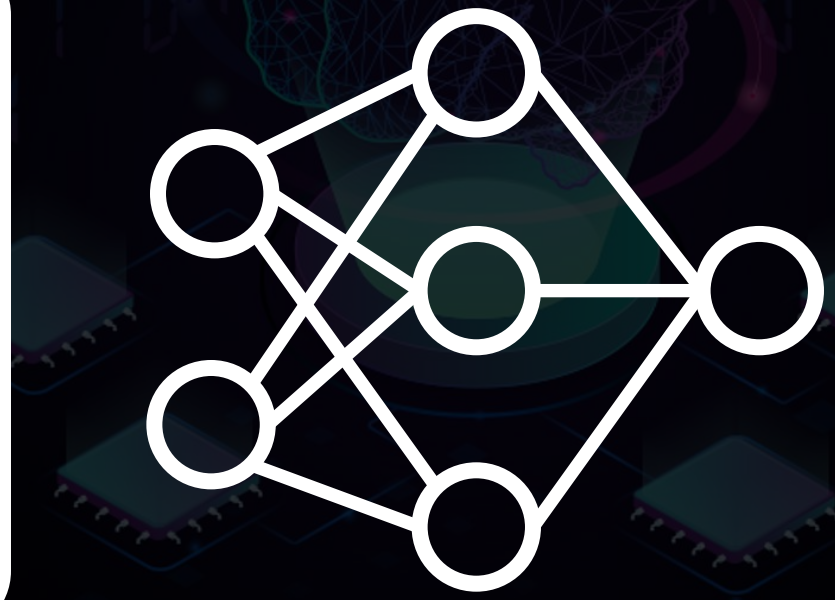
환경 → 에이전트 → 상태 → **행동** → 정책 → 보상 → 가치함수



강화학습 용어

환경 → 에이전트 → 상태 → 행동 → **정책** → 보상 → 가치함수

상태
(ex. 현위치)



행동
(ex. 움직임)

강화학습 용어

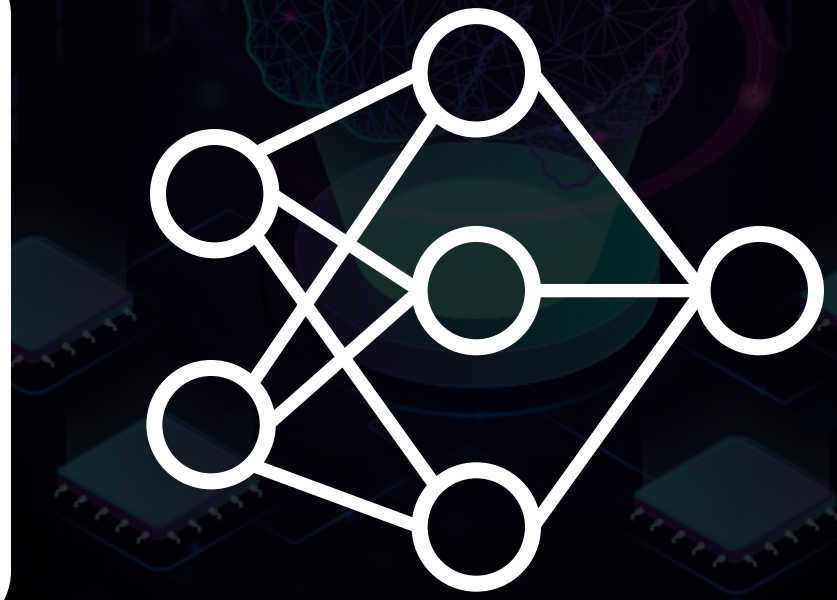
환경 → 에이전트 → 상태 → 행동 → 정책 → **보상** → 가치함수



강화학습 용어

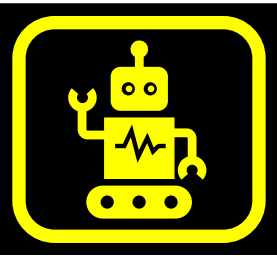
환경 → 에이전트 → 상태 → 행동 → 정책 → 보상 → 가치함수

상태
(ex. 현위치)



점수

강화학습 용어(정리)

환경			
(1,4)	(2,4)	(3,4)	목표
(1,3)	(2,3)	(3,3)	(4,3)
(1,2)	(2,2)	(3,2)	(4,2)
	(2,1)	(3,1)	(4,1)

1. 에이전트(로봇)의 상태(현위치) 파악

위치는? **(1,1)**

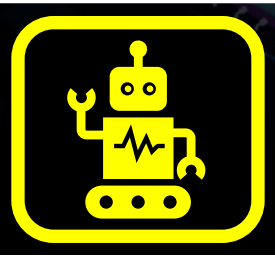
강화학습 용어(정리)



2. 정책 신경망으로
행동을 결정

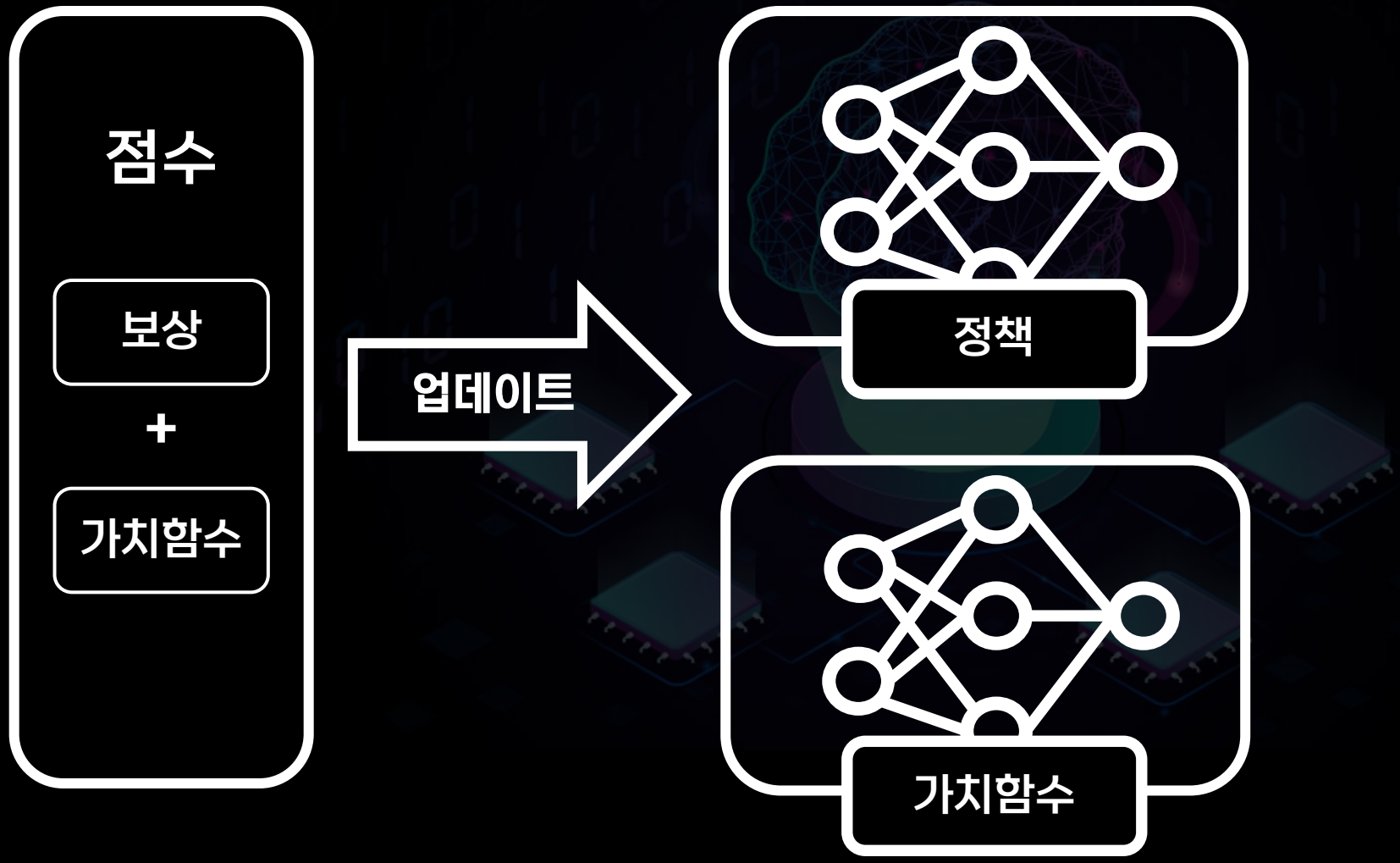
강화학습 용어(정리)

환경

(1,4)	(2,4)	(3,4)	목표
(1,3)	(2,3)	(3,3)	(4,3)
(1,2)	(2,2)	(3,2)	(4,2)
(1,1)		(3,1)	(4,1)

3. 행동을 수행하고,
보상과 가치함수 계산

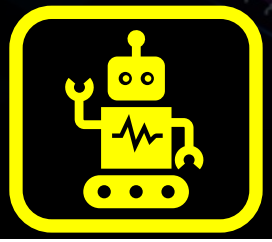
강화학습 용어(정리)



4. 정책과 가치함수
신경망 업데이트

강화학습 용어(정리)

환경

(1,4)	(2,4)	(3,4)	목표
(1,3)	(2,3)	(3,3)	(4,3)
(1,2)	(2,2)	(3,2)	(4,2)
(1,1)		(3,1)	(4,1)

로봇의 현위치 파악

↓

정책 신경망으로
행동을 결정

↓

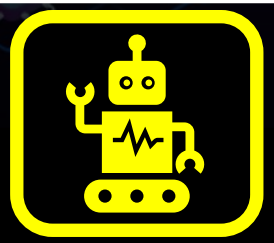
행동을 수행하고,
보상과 가치함수 계산

↓

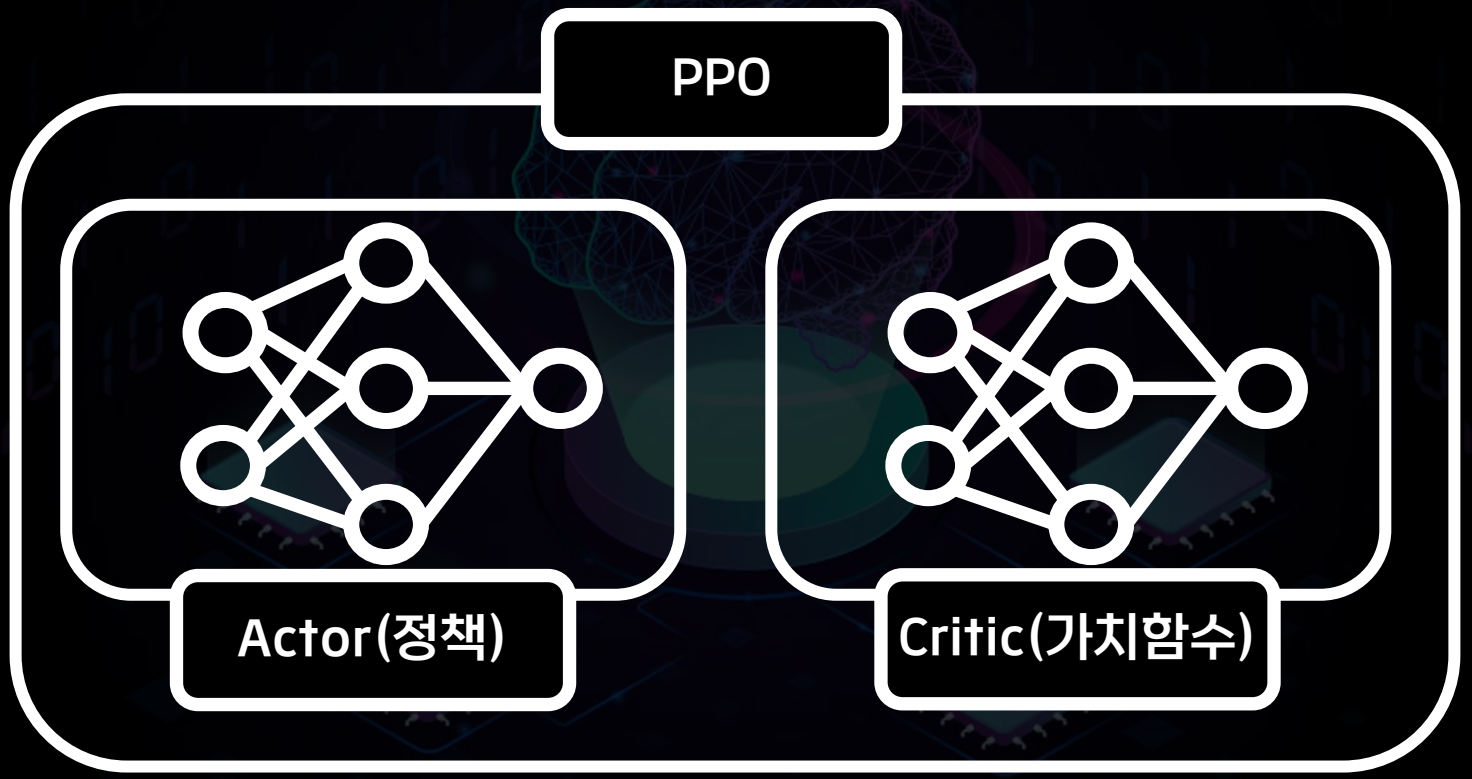
정책과 가치함수
신경망 업데이트

강화학습 용어(정리)

환경

(1,4)	(2,4)	(3,4)	목표
(1,3)	(2,3)	(3,3)	(4,3)
(1,2)	(2,2)	(3,2)	(4,2)
(1,1)		(3,1)	(4,1)

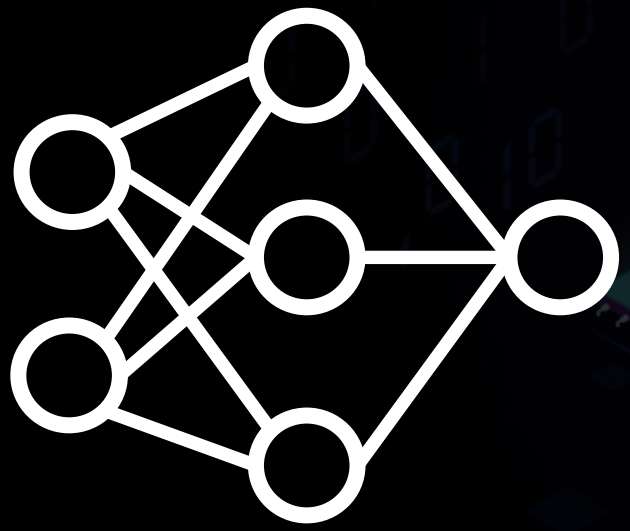
하이퍼파라미터, 병렬학습



PPO 알고리즘의 Actor-Critic 구조

하이퍼파라미터, 병렬학습

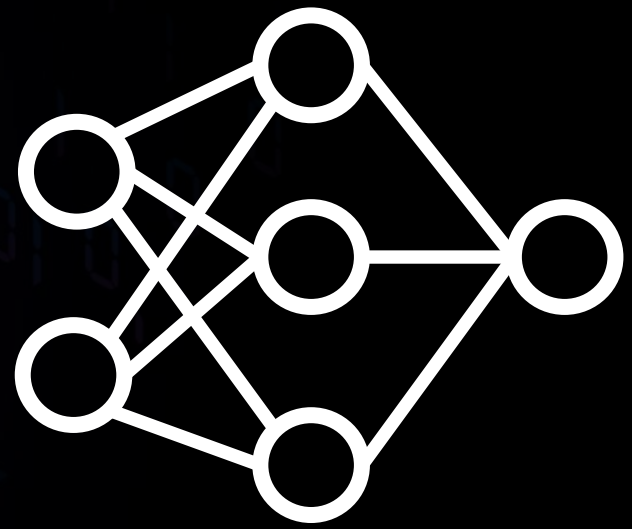
하이퍼파라미터



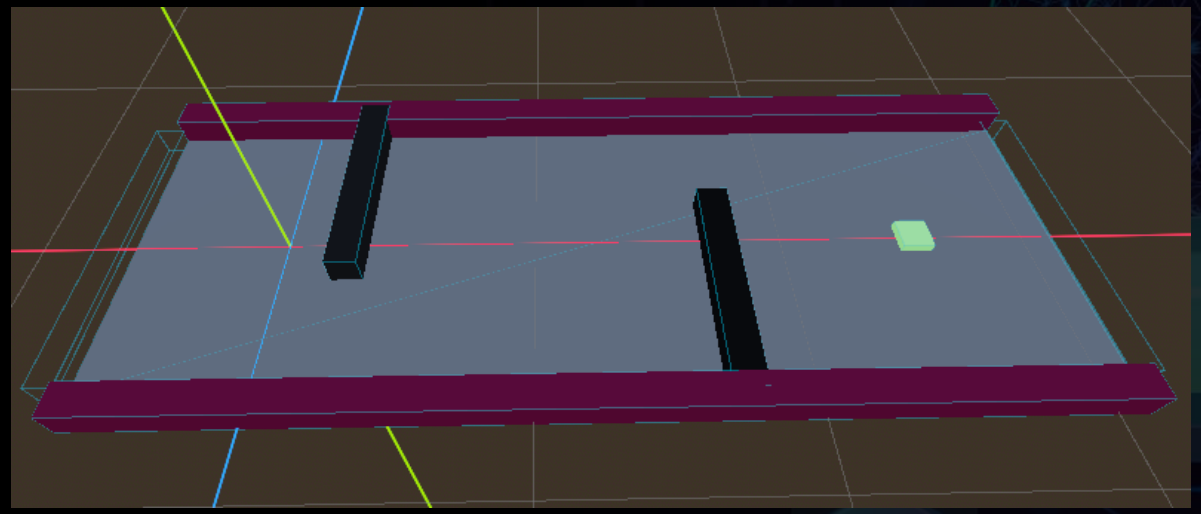
논문 or 예시

하이퍼파라미터, 병렬학습

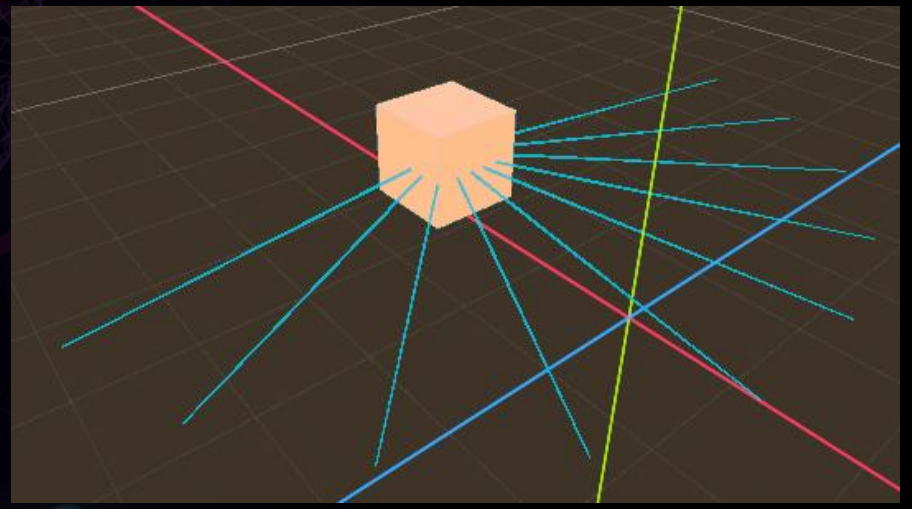
환경	환경	환경	환경
환경	환경	환경	환경
환경	환경	환경	환경
환경	환경	환경	환경



강화학습 예시

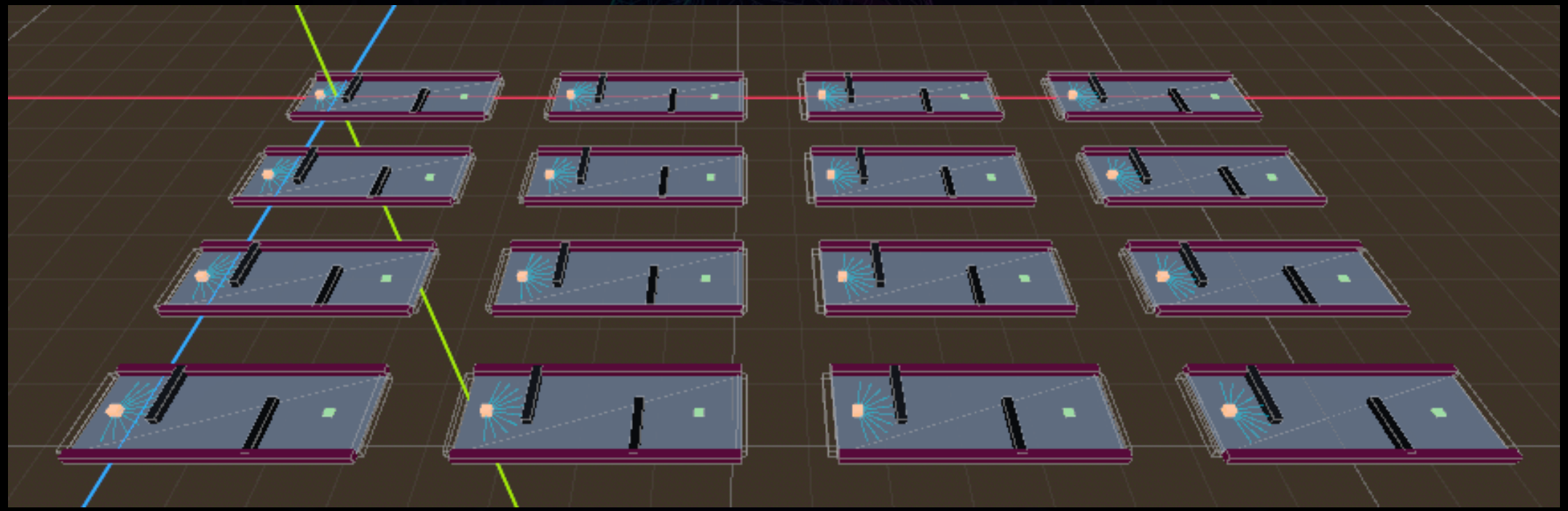


환경



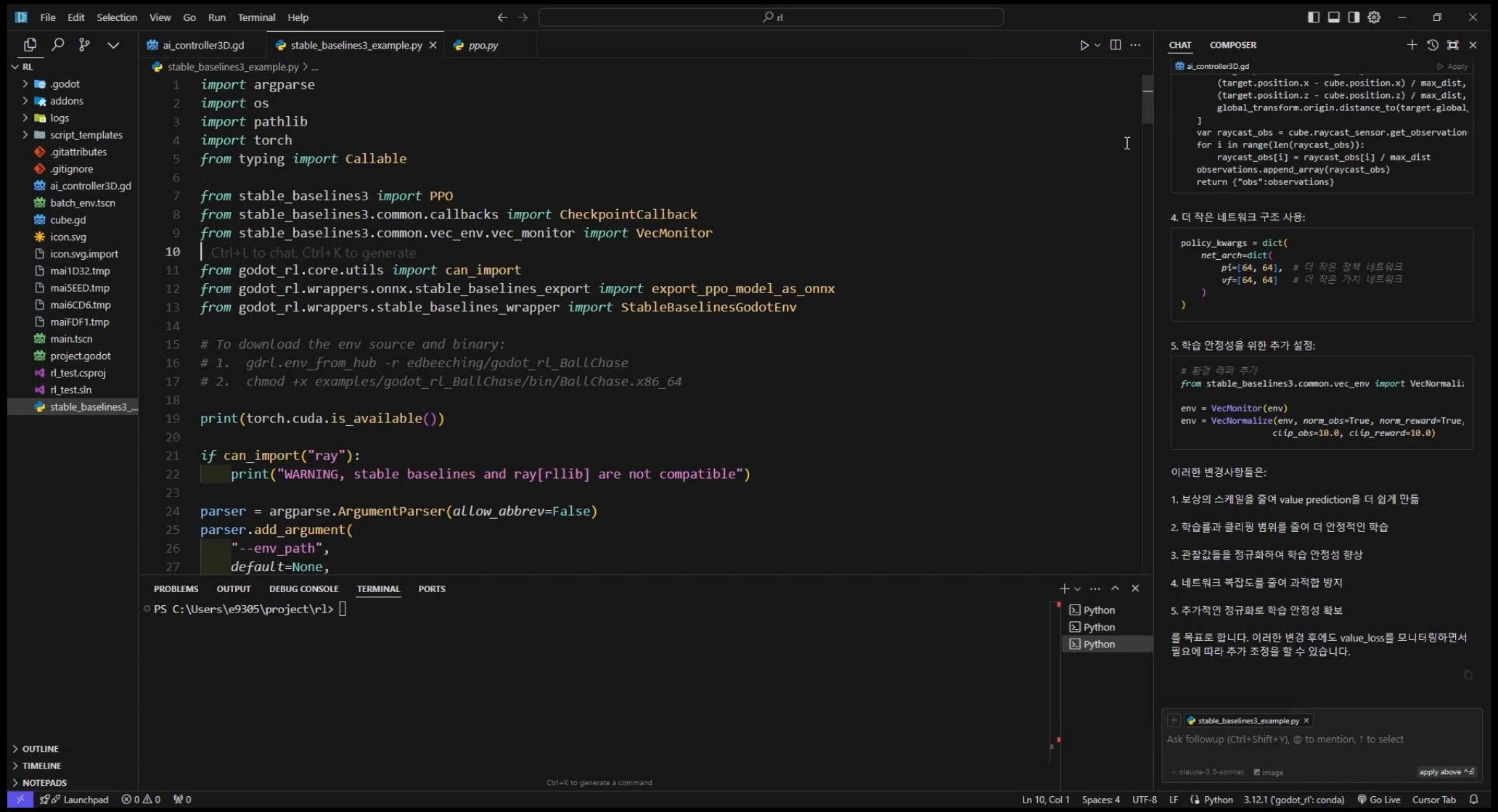
에이전트(로봇청소기)

강화학습 예시

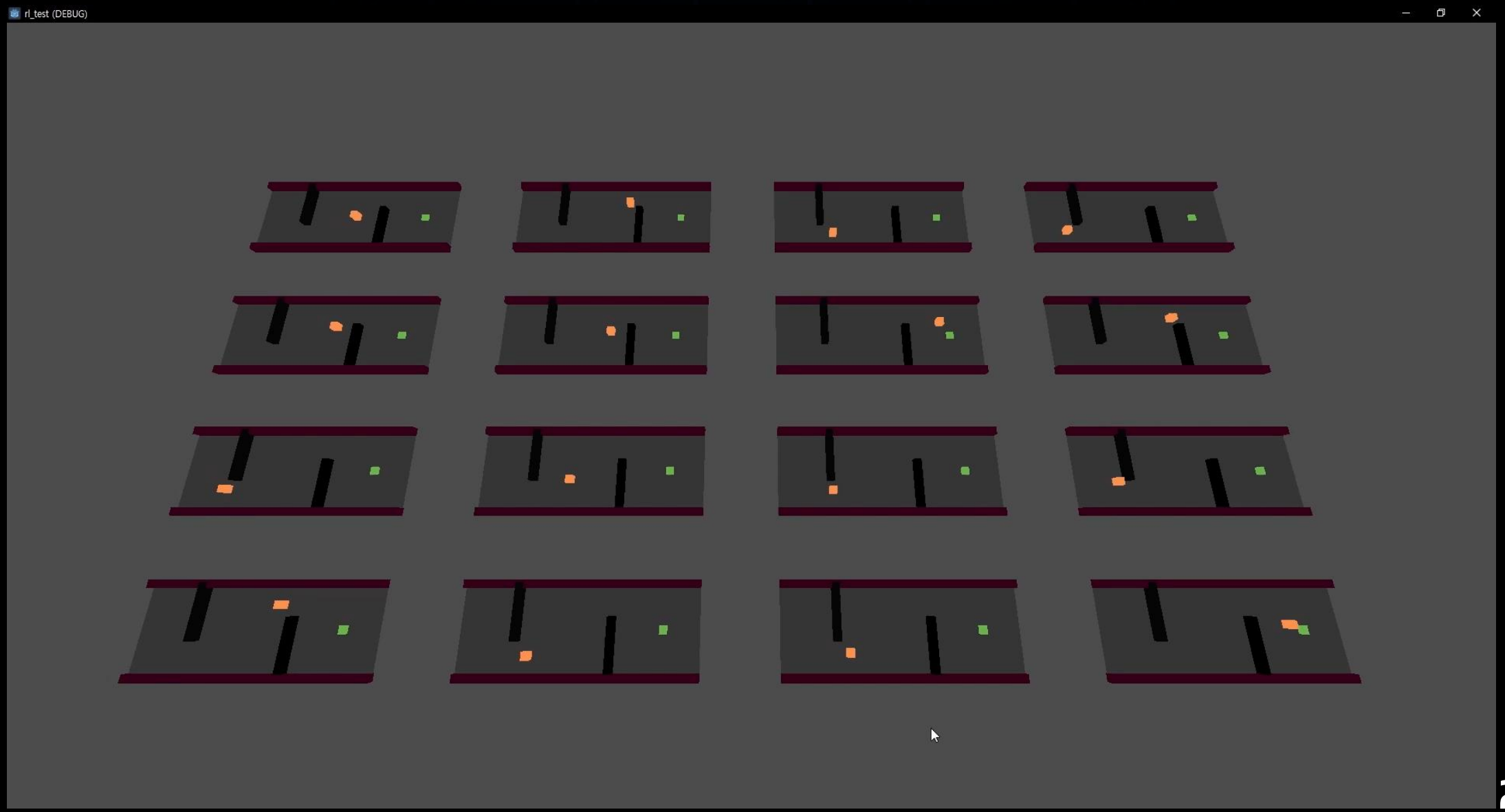


병렬학습

강화학습 예시


File Edit Selection View Go Run Terminal Help
ai_controller3D.gd
stable_baselines3_example.py
ppo.py
RL
.godot
addons
logs
script_templates
.gitattributes
.gitignore
ai_controller3D.gd
batch_env.tscn
cube.gd
icon.svg
icon.svg.import
main.tscn
project.godot
rl_test.csproj
rl_test.sln
import argparse
import os
import pathlib
import torch
from typing import Callable
from stable_baselines3 import PPO
from stable_baselines3.common.callbacks import CheckpointCallback
from stable_baselines3.common.vec_env.vec_monitor import VecMonitor
Ctrl+L to chat, Ctrl+K to generate
from godot_rl.core.utils import can_import
from godot_rl.wrappers.onnx.stable_baselines_export import export_ppo_model_as_onnx
from godot_rl.wrappers.stable_baselines_wrapper import StableBaselinesGodotEnv
# To download the env source and binary:
# 1. gdrl.env_from_hub -r edbeeching/godot_rl_BallChase
# 2. chmod +x examples/godot_rl_BallChase/bin/BallChase.x86_64
print(torch.cuda.is_available())
if can_import("ray"):
print("WARNING, stable baselines and ray[rllib] are not compatible")
parser = argparse.ArgumentParser(allow_abbrev=False)
parser.add_argument(
"--env_path",
default=None,
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
PS C:\Users\e9305\project\rl>
Python
Python
Python
CHAT COMPOSER
4. 더 작은 네트워크 구조 사용:
policy_kwargs = dict(
net_arch=dict(
pi=[64, 64], # 더 작은 정책 네트워크
vf=[64, 64] # 더 작은 가치 네트워크
)
)
5. 학습 안정성을 위한 추가 설정:
# 환경 래퍼 추가
from stable_baselines3.common.vec_env import VecNormali
env = VecMonitor(env)
env = VecNormalize(env, norm_obs=True, norm_reward=True,
clip_obs=10.0, clip_reward=10.0)
이러한 변경사항들은:
1. 보상의 스케일을 줄여 value prediction을 더 쉽게 만듦
2. 학습률과 클리핑 범위를 줄여 더 안정적인 학습
3. 관찰값들을 정규화하여 학습 안정성 향상
4. 네트워크 복잡도를 줄여 과적합 방지
5. 추가적인 정규화로 학습 안정성 확보
를 목표로 합니다. 이러한 변경 후에도 value_loss를 모니터링하면서 필요에 따라 추가 조정을 할 수 있습니다.
Ask followup (Ctrl+Shift+Y), @ to mention, ↑ to select
apply above
OUTLINE
TIMELINE
NOTEPADS
Launchpad
Ln 10, Col 1
Spaces: 4
UTF-8
LF
Python 3.12.1 ('godot_rl': conda)
Go Live
Cursor Tab

강화학습 예시



사업단에서 고려할점

- 환경과 에이전트의 정의(요구사항)



사업단에서 고려할점

사업단에서 고려할점

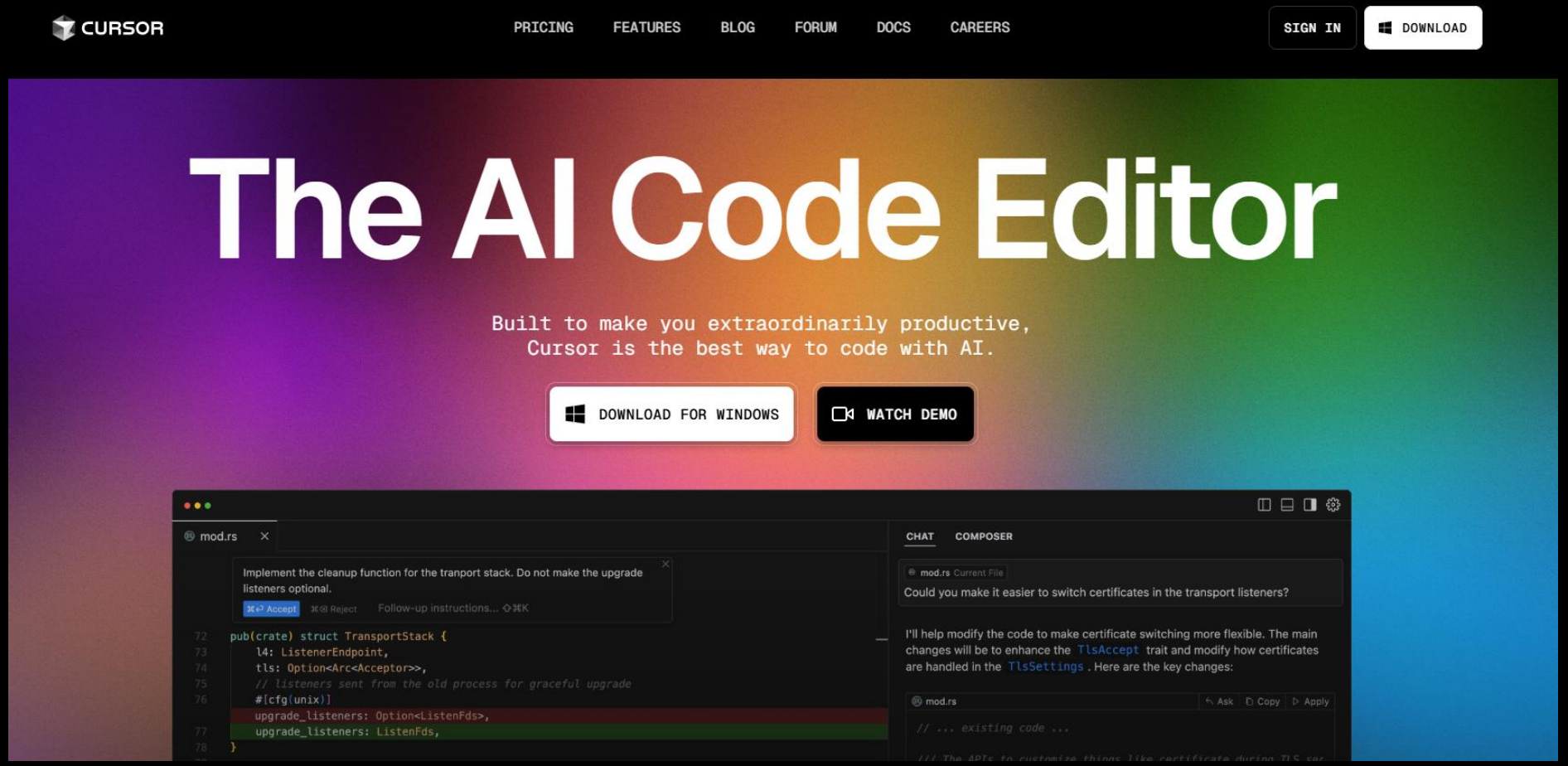
- 시뮬레이션과 현실과 괴리(Domain Gap)





인공지능 툴 사용법

- Cursor 설치

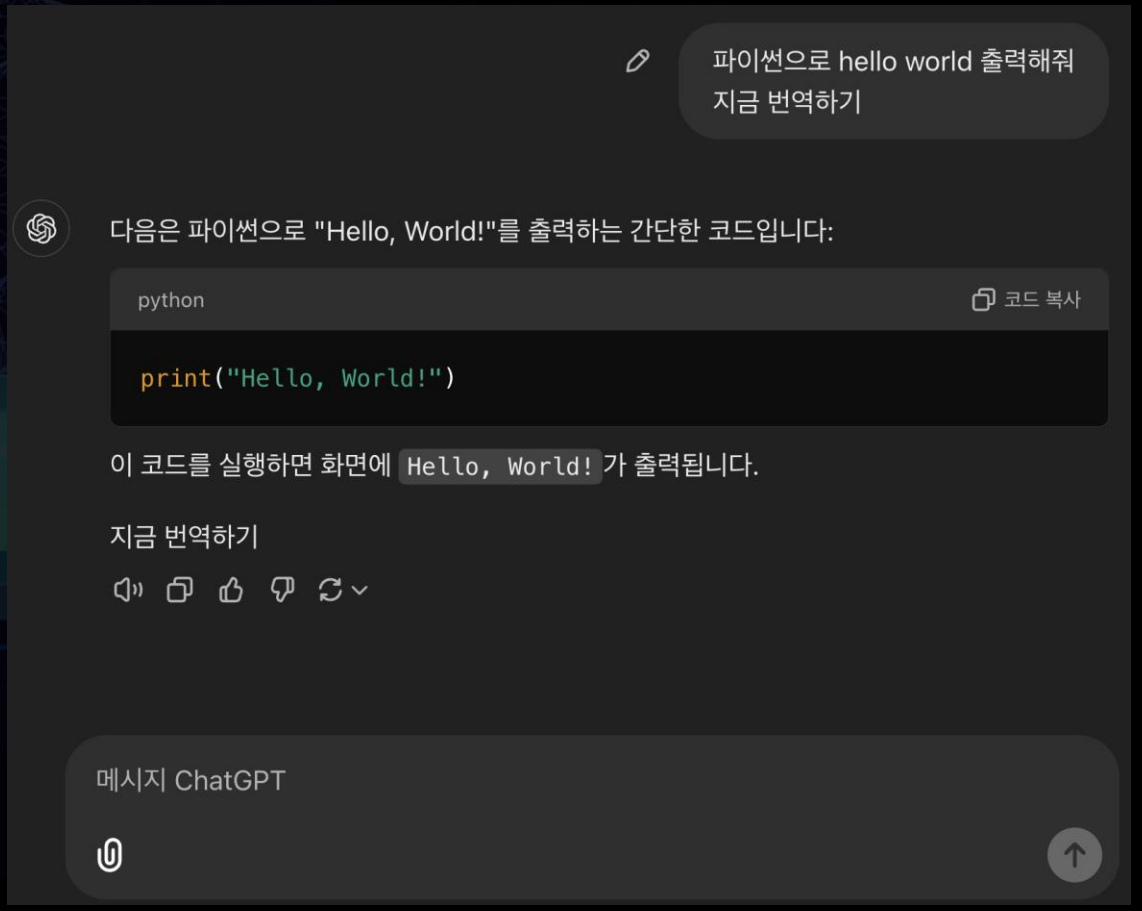


https://www.cursor.com/

인공지능 툴 사용법

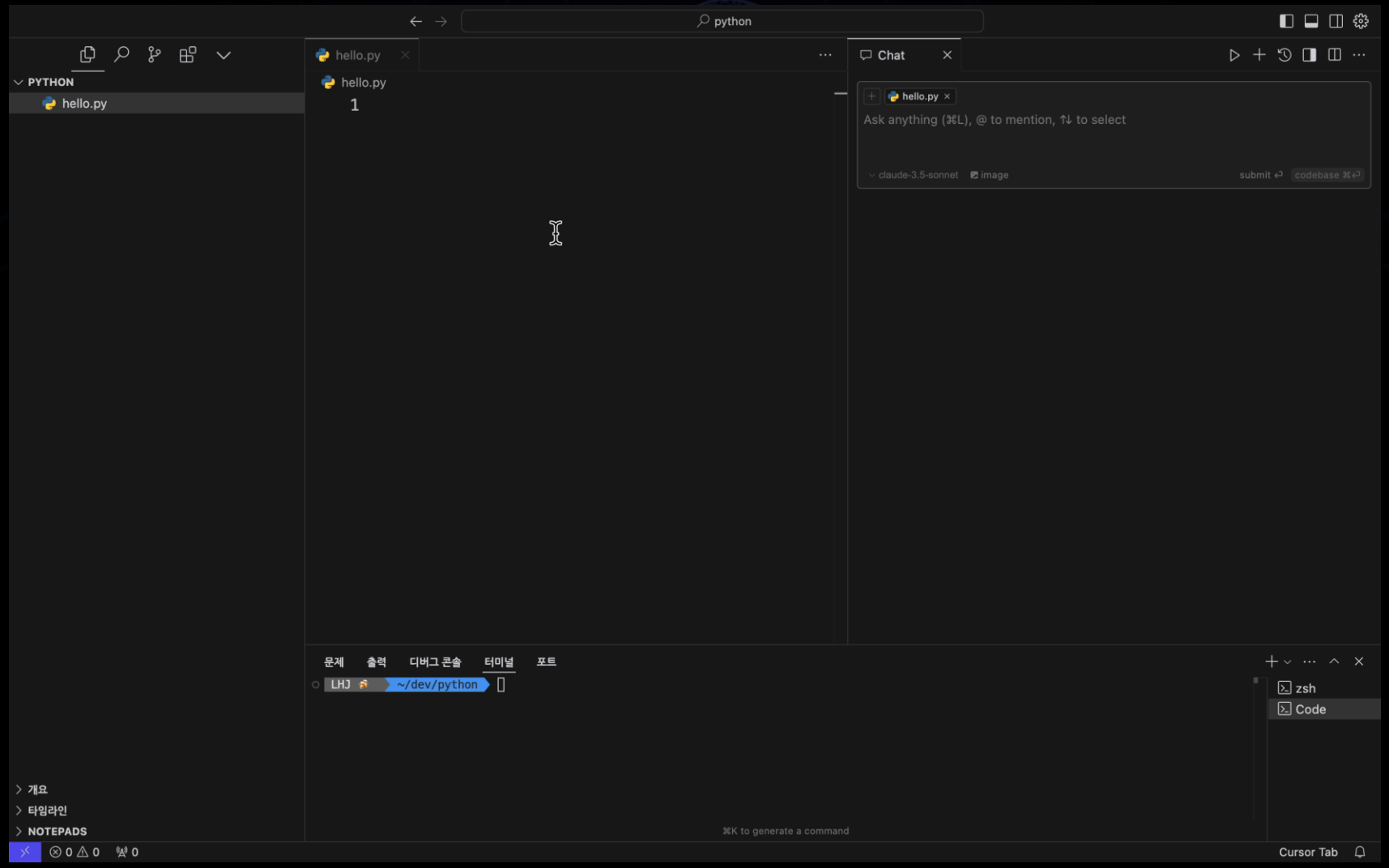


Visual Studio Code



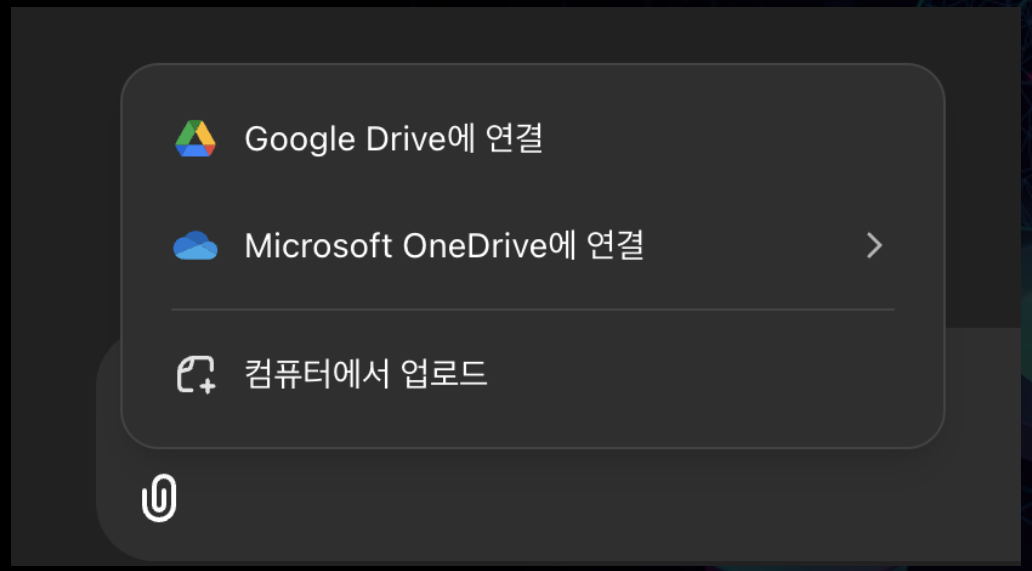
인공지능 툴 사용법

- Cursor에 채팅해서 코드 작성



인공지능 툴 사용법

- RAG 예시



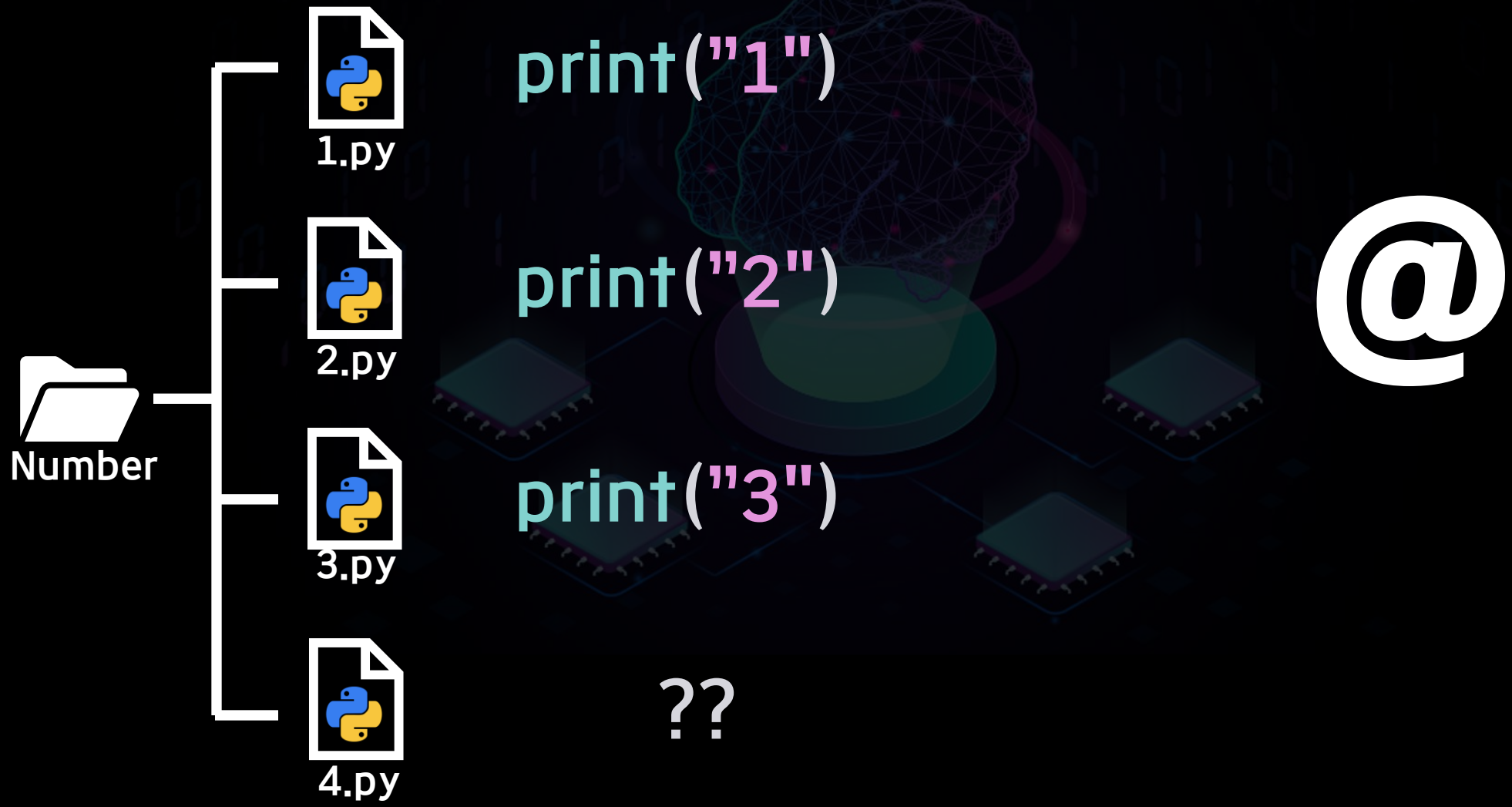
PDF



Web 검색

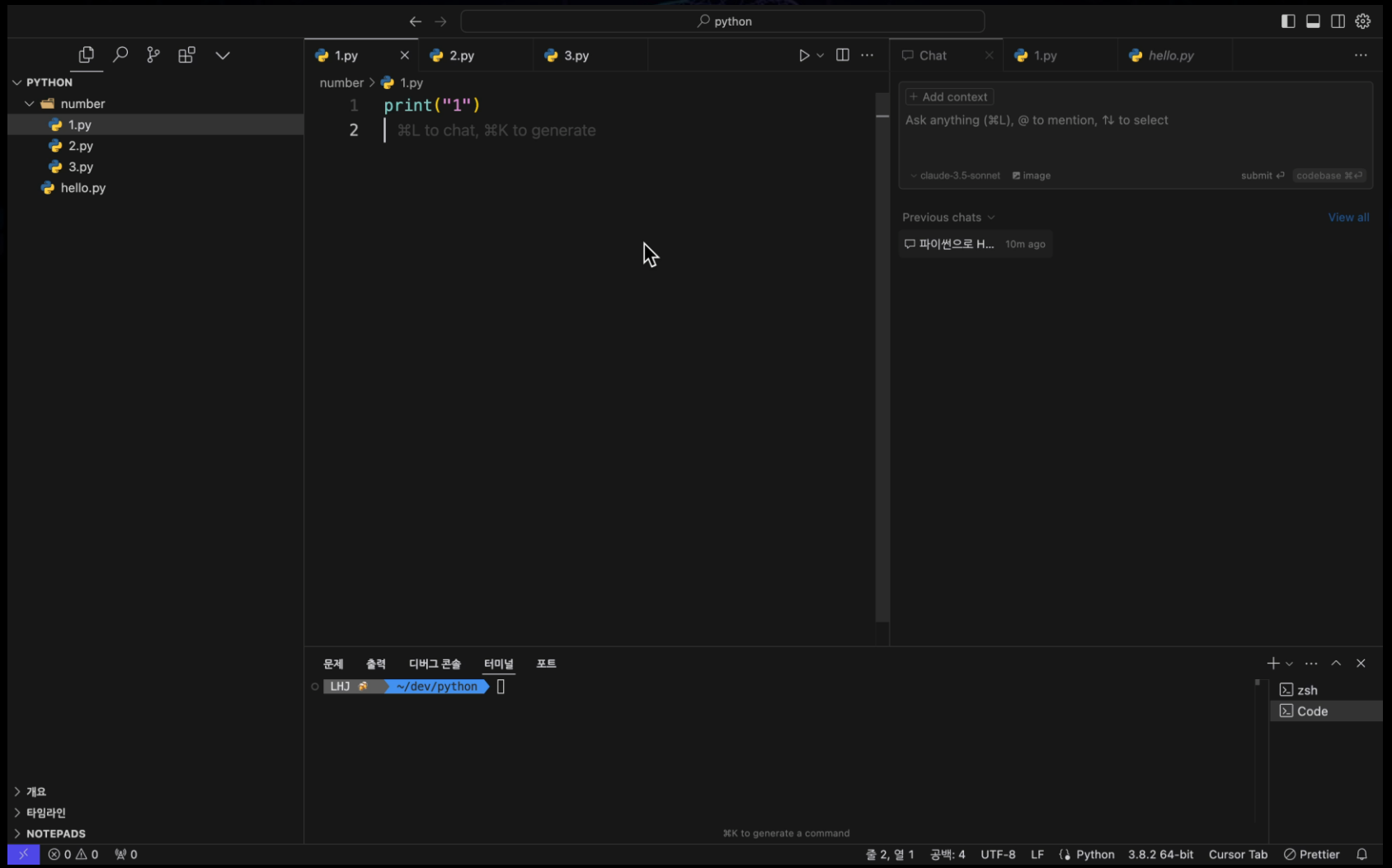
인공지능 툴 사용법

- Cursor 채팅해서 코드 작성 + RAG



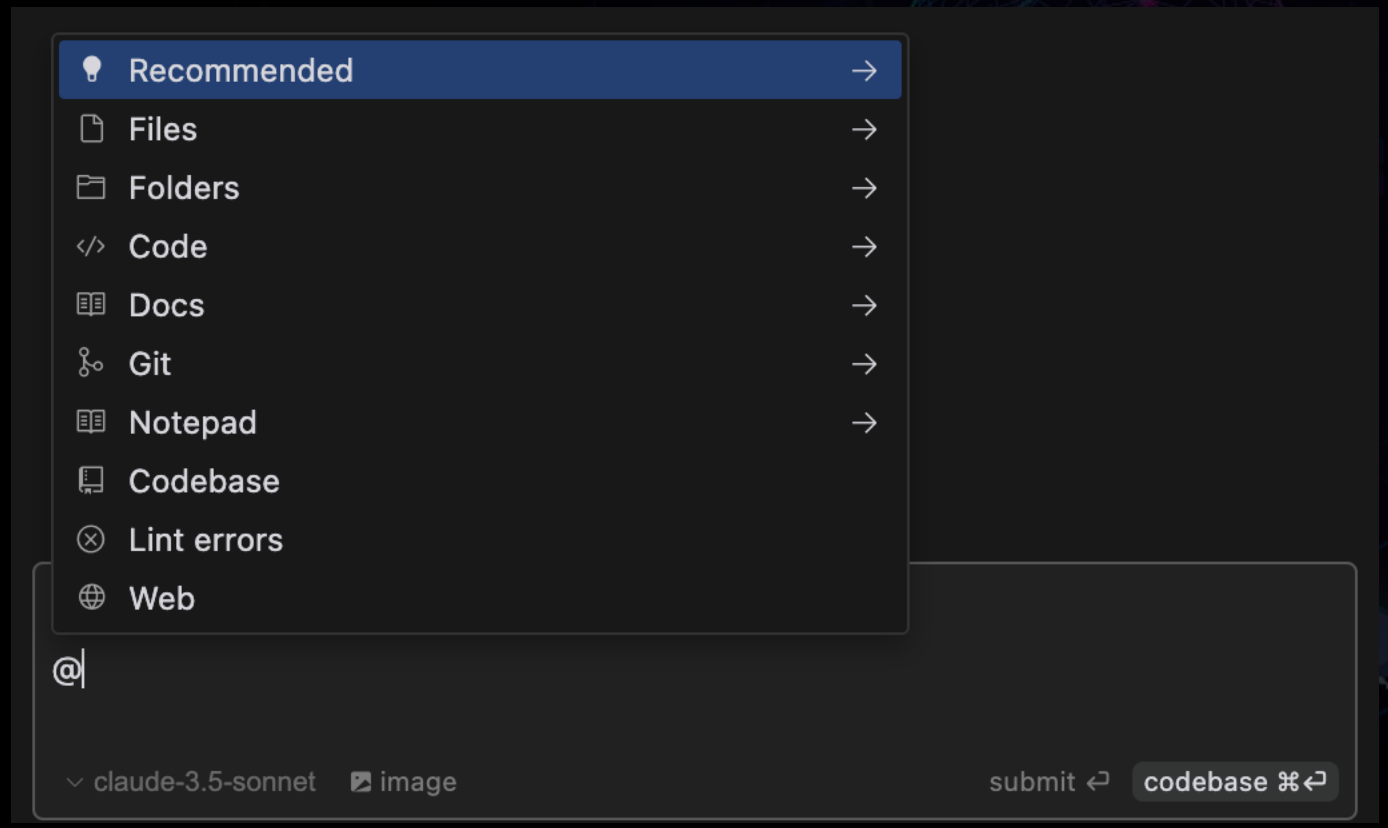
인공지능 툴 사용법

- Cursor 채팅해서 코드 작성 + RAG



인공지능 툴 사용법

- Cursor 채팅해서 코드 작성 + RAG



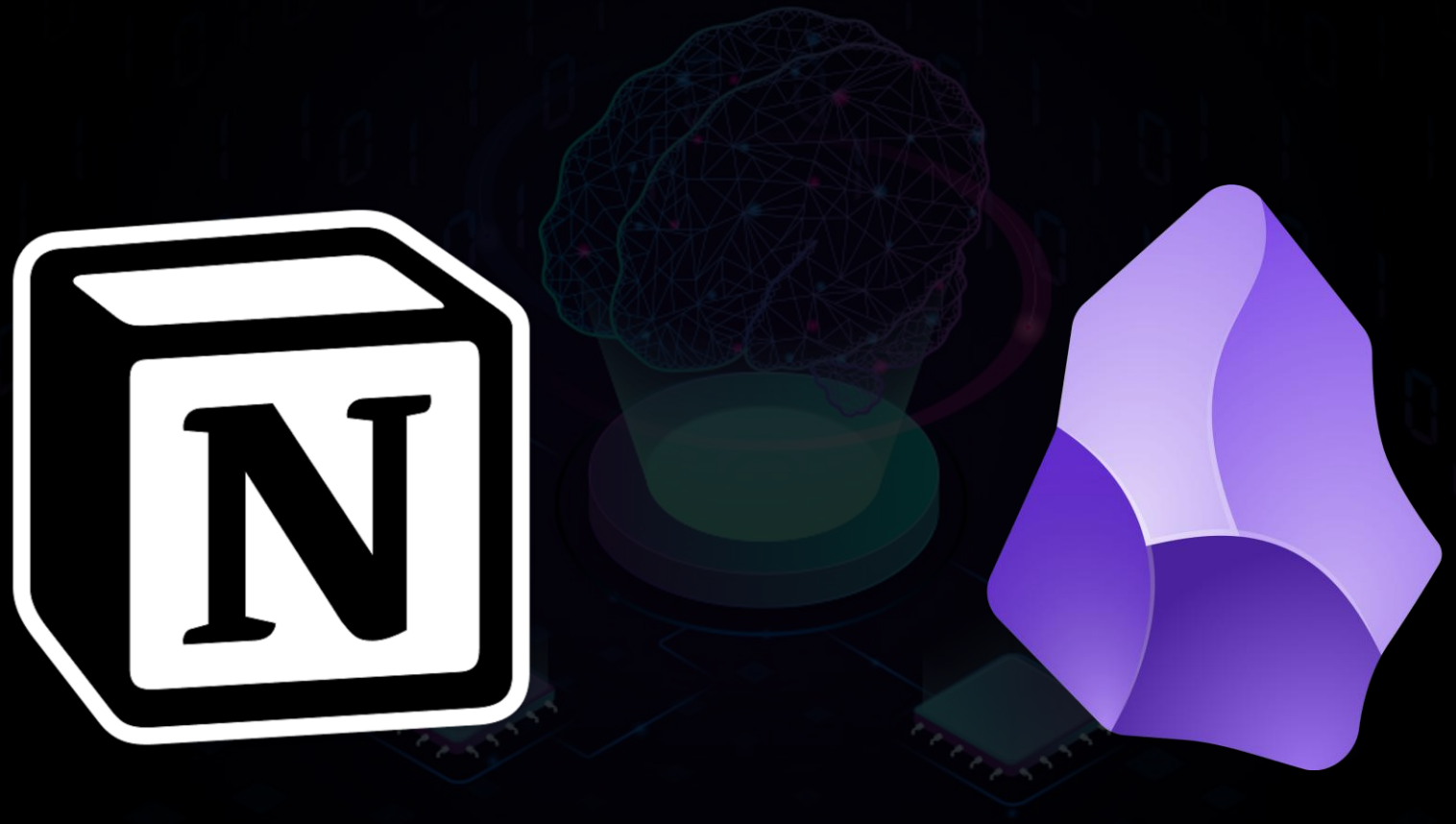
파일

폴더

Docs
(공식문서)

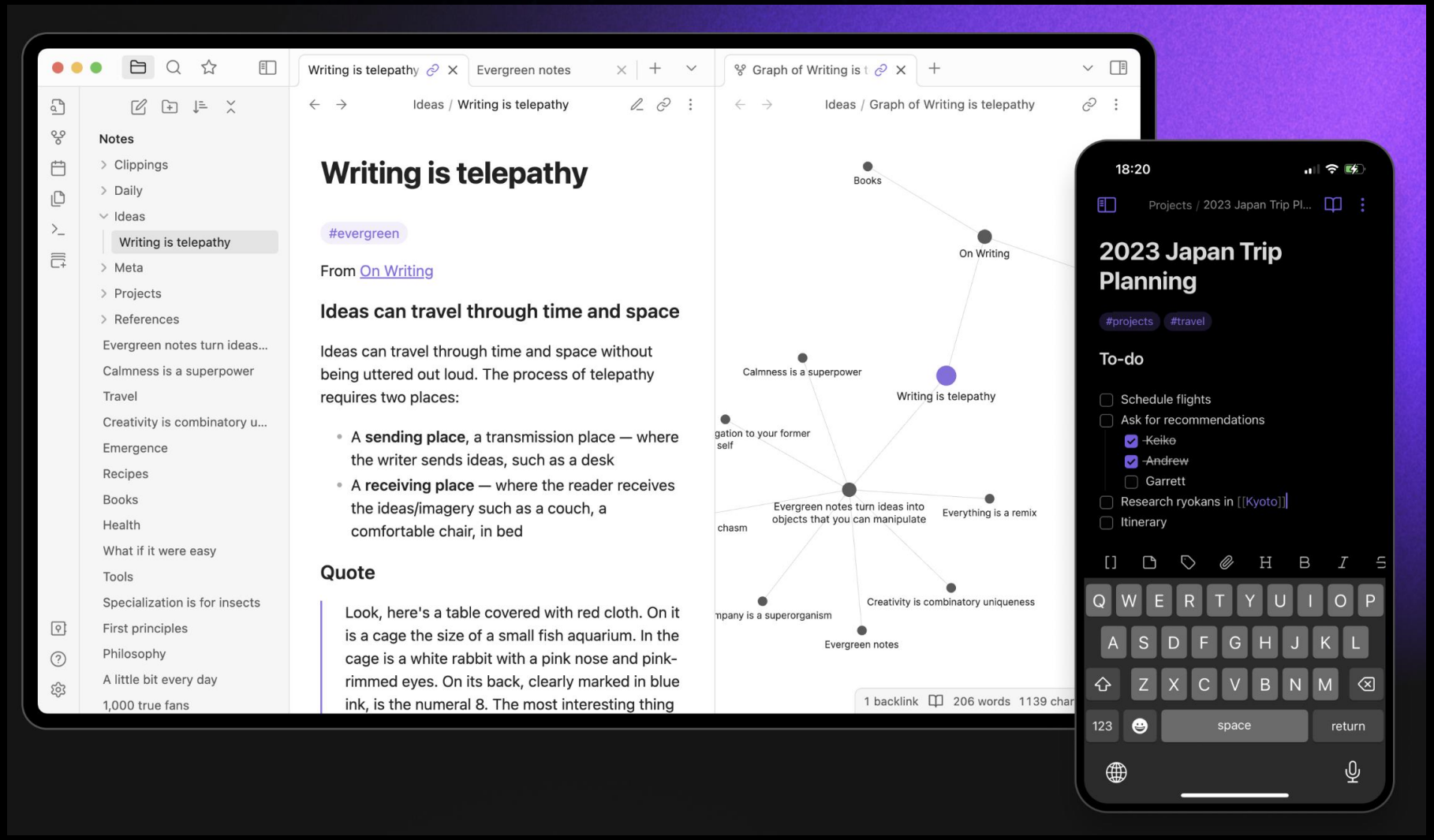
인공지능 툴 사용법

- Cursor+Obsidian



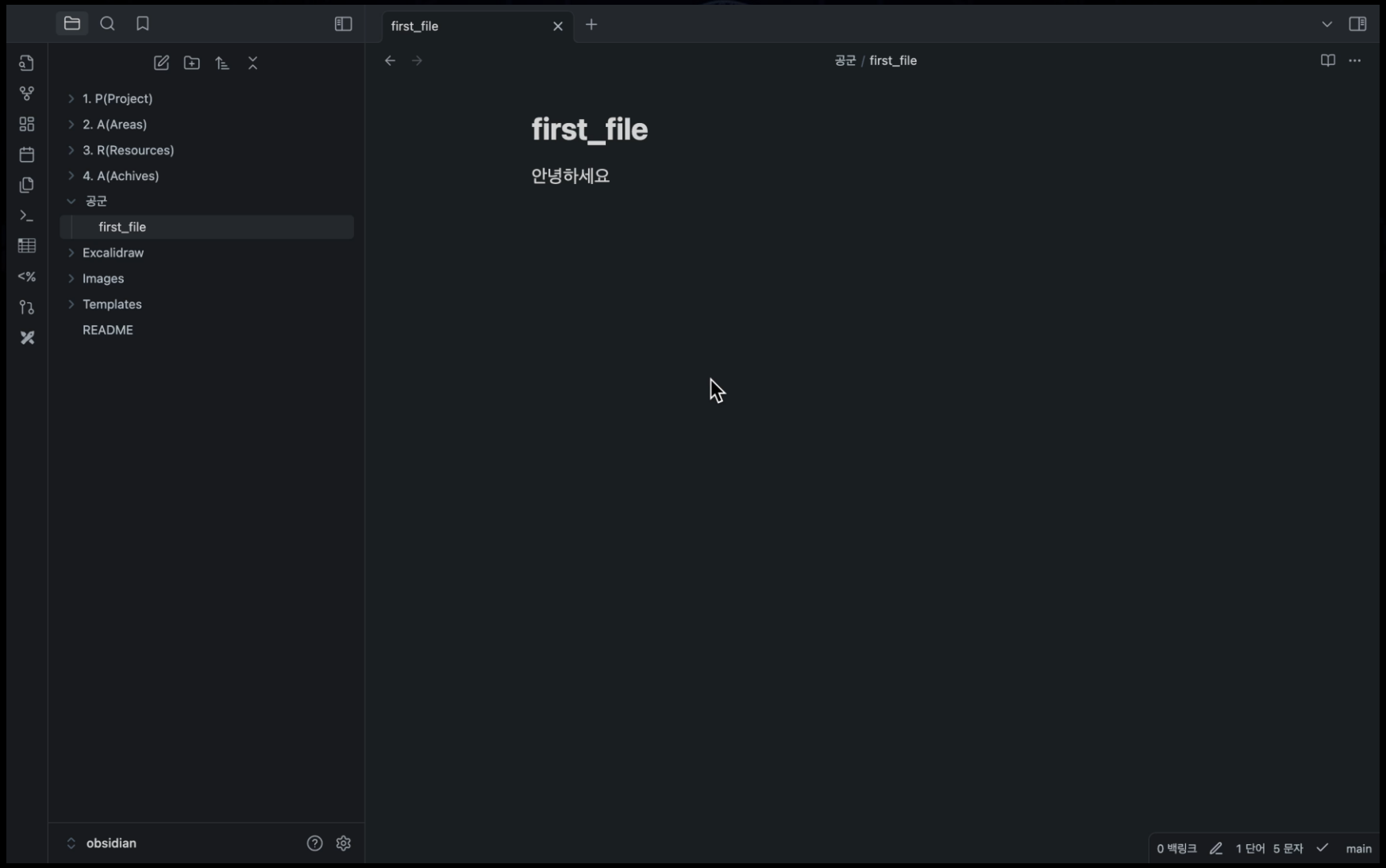
인공지능 툴 사용법

- Cursor+Obsidian



인공지능 툴 사용법

- Cursor+Obsidian



인공지능 툴 사용법

- Cursor+Obsidian(일일 업무 보고서 작성)

Obsidian에서 관련된 **이전 보고서들을 검색**

↓

Cursor에서 '**@**' 기능으로 해당 자료들을 **참조**

↓

AI에게 "이 내용들을 참고해서 더 간결하고 명확한 보고서 형식으로 만들어줘"라고 **요청**

↓

AI가 제안한 내용을 검토하고 수정하여 **최종 문서 완성**


대한민국공군
REPUBLIC OF KOREA AIR FORCE
2024
공군 창의·혁신
아이디어 공모
해커톤
제6회
공군 창의·혁신
아이디어 공모
해커톤
자유공모
대 상
(공군참모총장상)
₩5,000,000원
서울특별시
IBK기업은행
SW KOREA


감사합니다.