AI+ 软件研发数字峰会
AI+ software Development Digital summit

AI驱动软件研发
全面进入数字化时代

中国·北京 08.18-19

AI+
software
Development
Digital
summit

OpenRL: A Unified Reinforcement Learning Framework

黄世宇 @ 第四范式

科技生态圈峰会 + 深度研习 ——1000 + 技术团队的共同选择

K+峰会

2023K+
全球软件研发行业创新峰会
上海站
会议时间 | 06.09-10

2023K+
全球软件研发行业创新峰会
北京站
会议时间 | 07.21-22

2024K+
全球软件研发行业创新峰会
深圳站
会议时间 | 05.17-18

K+峰会详情

AiDD峰会

会议时间 | 08.18-19
AiDD AI+软件研发数字峰会
北京站

会议时间 | 11.17-18
AiDD AI+软件研发数字峰会
深圳站

AiDD峰会详情

演讲嘉宾

黄世宇

第四范式强化学习科学家，开源强化学习OpenRL Lab负责人

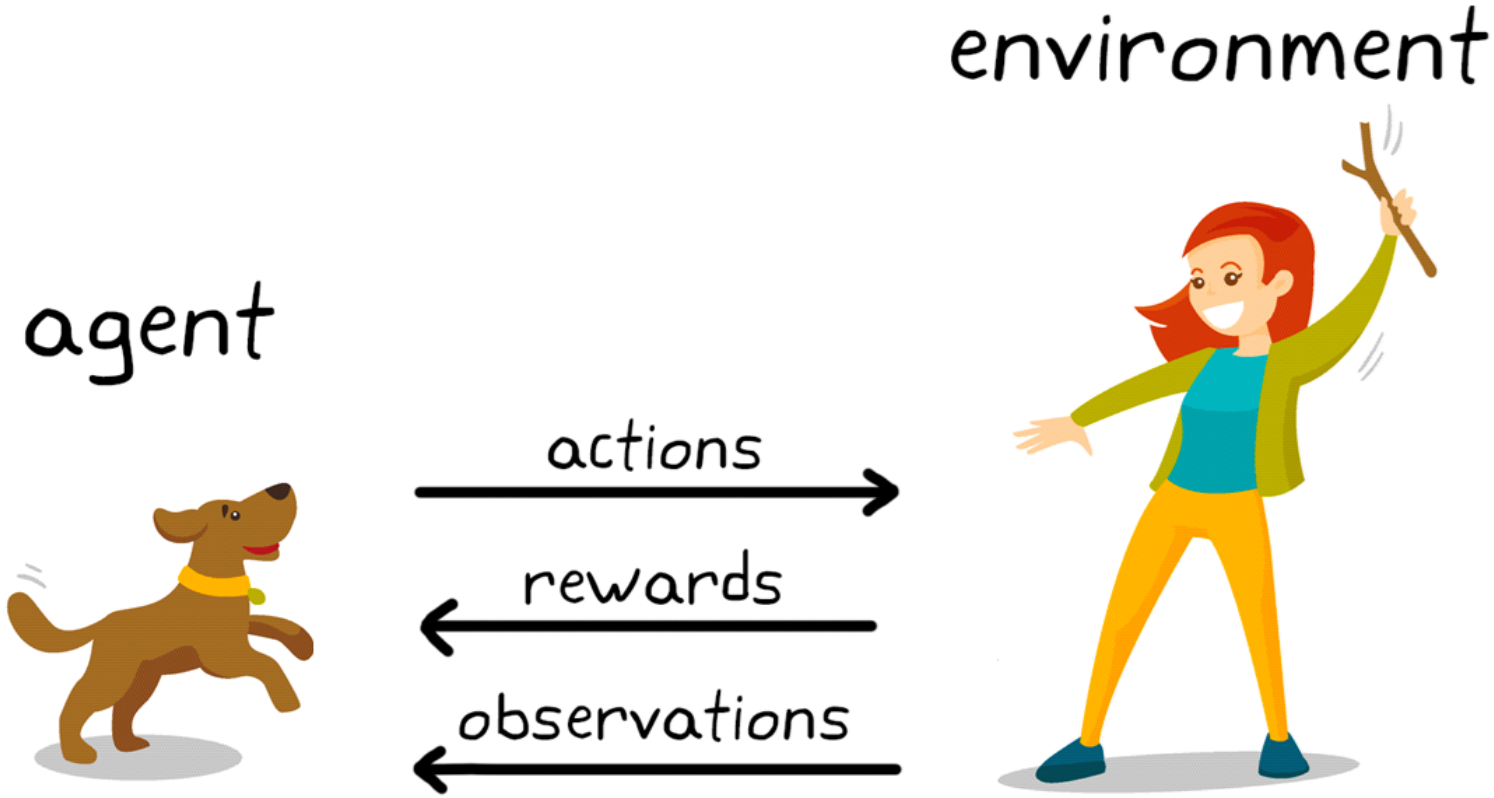
本科与博士均毕业于清华大学计算机系，导师是朱军和陈挺教授，本科期间在CMU交换，导师为Deva Ramanan教授。主要研究方向为强化学习，多智能体强化学习，分布式强化学习。曾在ICLR、CVPR、AAAI、NeurIPS, Nature Machine Intelligence, ICML, AAMAS, Pattern Recognition等会议和期刊发表多篇学术论文。其领导开发的TiZero谷歌足球游戏智能体曾在及第平台上取得排名第一的成绩。黄世宇也曾在腾讯AI Lab、华为诺亚、商汤、瑞莱智慧等工作。

目录
CONTENTS

PART 01
Introduction & Motivation

What is Reinforcement Learning?

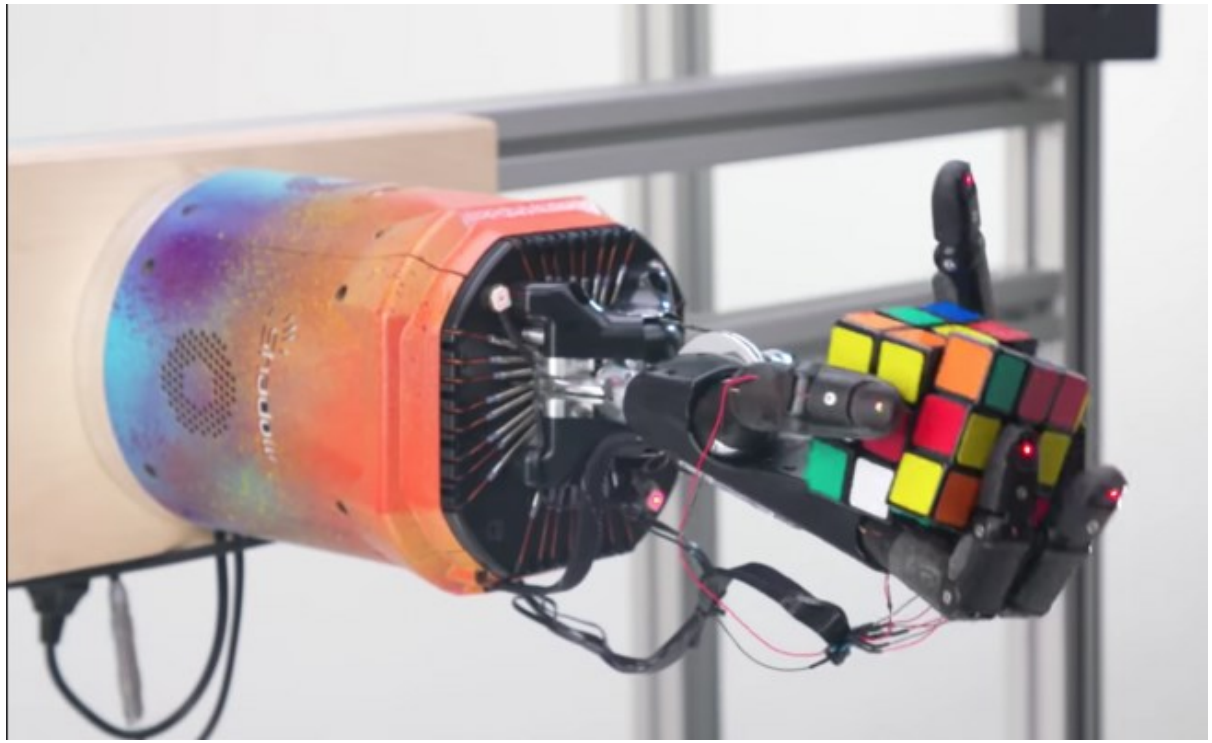
- **Goal of RL**: Artificial General Intelligence (**AGI**)



Reinforcement learning in dog training.

What else?

- **Robotics**



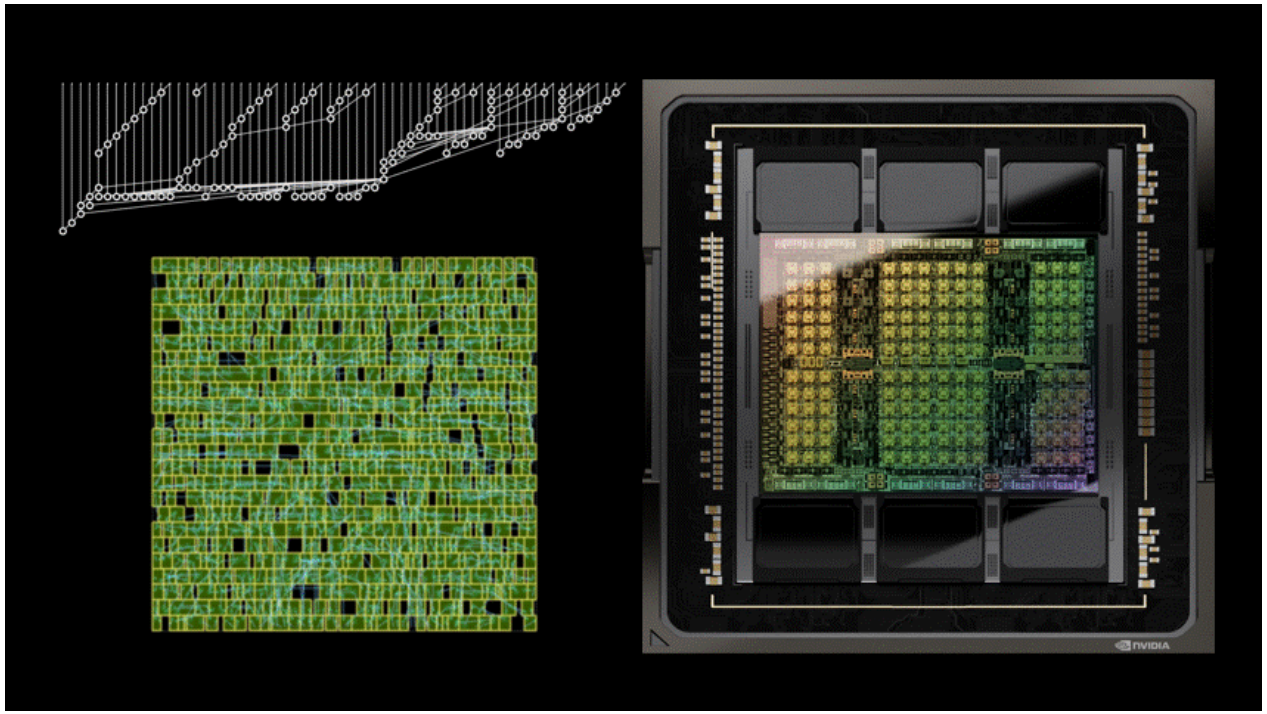
OpenAI 2019

➢ Autonomous Driving

CARLA 2017

What else?

- **Industrial Design**



PrefixRL 2022

➢ Quantitative Trading

FinRL 2020

What else?

- Chat Bot

Step 3

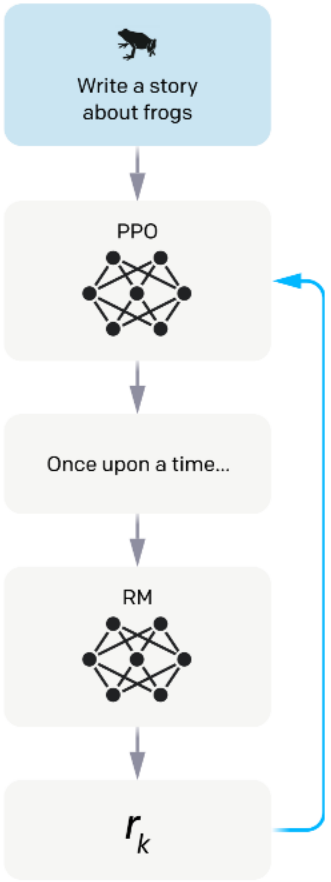
Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



What else?

- **Multi-agent RL**

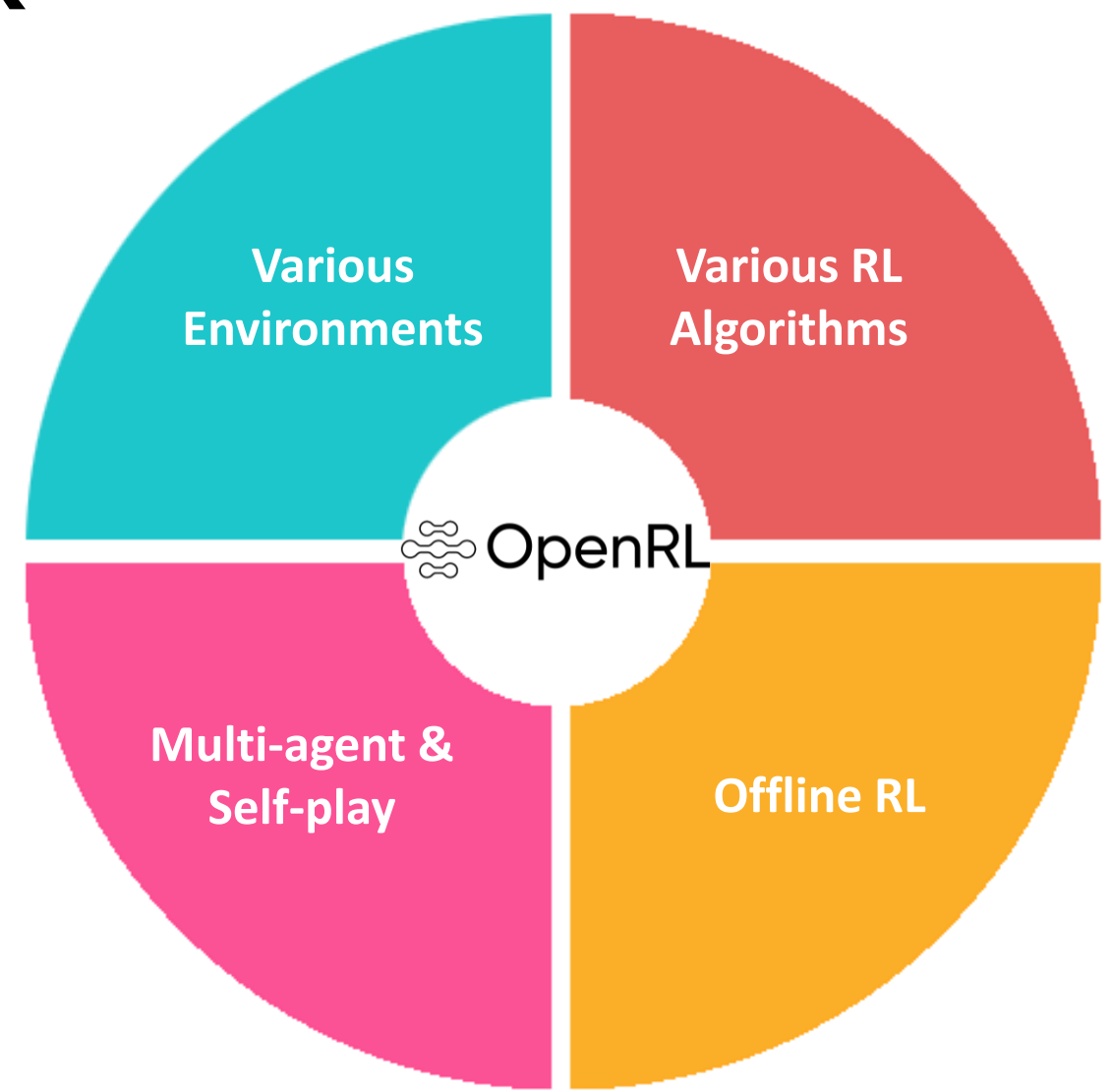
TiZero 2023

➢ **Competitive RL**

Honor of Kings Arena 2022

Do RL in a Unified Framework

```
env = make("env_name")
net = Net(env)
agent = Agent(net)
agent.train(total_time_steps=100)
obs, info = env.reset()
while True:
    action, _ = agent.act(obs)
    obs, r, done, info = env.step(action)
```



PART 02

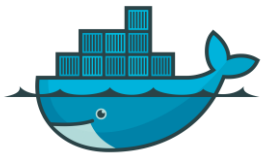
OpenRL: An Open-Souce RL Framework

Main Features of OpenRL

Friendly to beginners

```
pip install openrl
```

or



Google Colaboratory

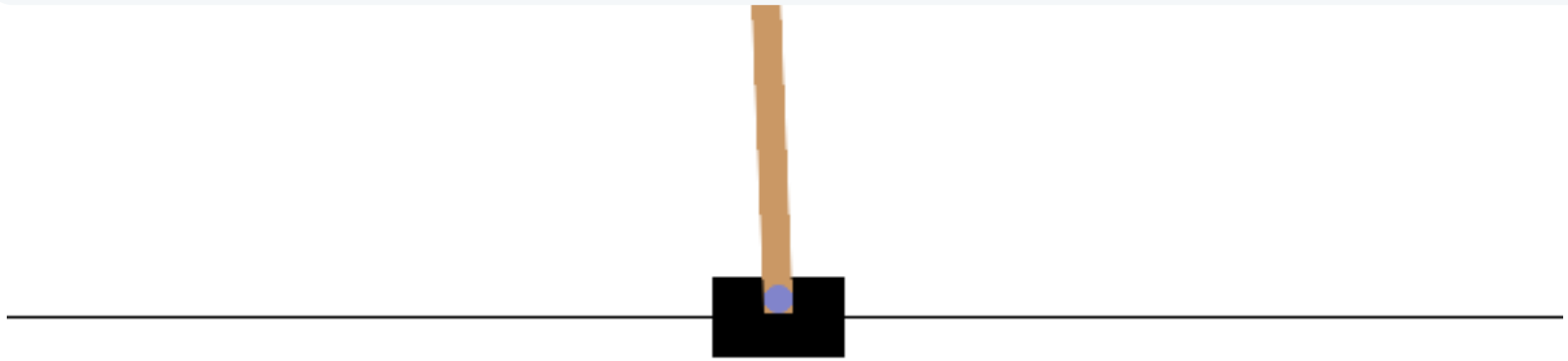
CO

```
docker pull openrllab/openrl
```

Main Features of OpenRL

Friendly to beginners

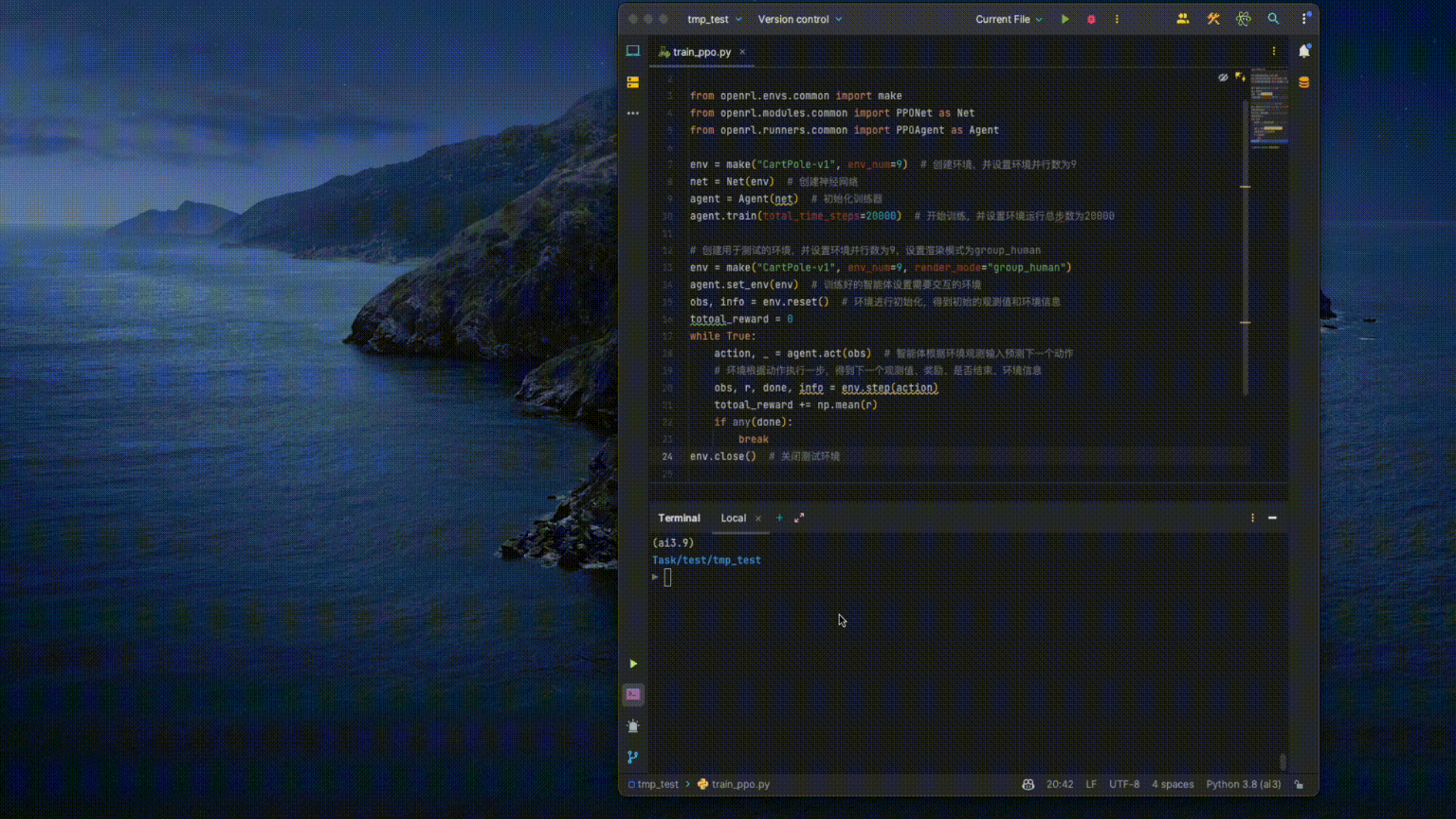
```
openrl --mode train --env CartPole-v1
```



Main Features of OpenRL

Friendly to beginners

```python
from openrl.envs.common import make
from openrl.modules.common import PPONet as Net
from openrl.runners.common import PPOAgent as Agent
env = make("CartPole-v1", env_num=9)
net = Net(env) # create the neural network
agent = Agent(net) # initialize the trainer
# start training, set total number of training steps to 20000
agent.train(total_time_steps=20000)
```


tmp_test
Version control
Current File
train_ppo.py
from openrl.envs.common import make
from openrl.modules.common import PPONet as Net
from openrl.runners.common import PPOAgent as Agent
env = make("CartPole-v1", env_num=9) # 创建环境，并设置环境并行数为9
net = Net(env) # 创建神经网络
agent = Agent(net) # 初始化训练器
agent.train(total_time_steps=20000) # 开始训练，并设置环境运行总步数为20000
# 创建用于测试的环境，并设置环境并行数为9，设置渲染模式为group_human
env = make("CartPole-v1", env_num=9, render_mode="group_human")
agent.set_env(env) # 训练好的智能体设置需要交互的环境
obs, info = env.reset() # 环境进行初始化，得到初始的观测值和环境信息
totoal_reward = 0
while True:
action, _ = agent.act(obs) # 智能体根据环境观测输入预测下一个动作
环境根据动作执行一步，得到下一个观测值、奖励、是否结束、环境信息
obs, r, done, info = env.step(action)
totoal_reward += np.mean(r)
if any(done):
break
env.close() # 关闭测试环境
Terminal
Local
(ai3.9)
Task/test/tmp_test
tmp_test
train_ppo.py
20:42
LF
UTF-8
4 spaces
Python 3.8 (ai3)

Main Features of OpenRL

Friendly to beginners

Documentation/中文文档

WELCOME TO OPENRL'S DOCUMENTATION!

OpenRL

中文文档 | GitHub

User Guide

- OpenRL Introduction
  - OpenRL Reinforcement Learning Framework
  - Citing OpenRL
- Quick Start Guide
  - Installation Instructions
  - Train Your First Agent
  - Multi-Agent Training
  - Train Natural Language Dialogue Task
- API Doc
  - Subpackages

欢迎来到 OPENRL 中文文档

OpenRL

English | GitHub

用户指南

- OpenRL 介绍
  - OpenRL 强化学习框架
  - Citing OpenRL
- 快速上手
  - 安装说明
  - 开始智能体训练
  - 训练多智能体强化学习算法
  - 训练自然语言对话任务
- API Doc
  - Subpackages

Tutorial

TRAIN YOUR FIRST AGENT

Training Environment

OpenRL provides users with a simple and easy-to-use way of using it. Here we take the CartPole environment as an example, to demonstrate how to use OpenRL for reinforcement learning training. Create a new file train_ppo.py and enter the following code:

```
# train_ppo.py
from openrl.envs.common import make
from openrl.modules.common import PPONet as Net
from openrl.runners.common import PPOAgent as Agent
env = make("CartPole-v1", env_num=9) # create environment, set environment parallelism to 9
net = Net(env) # create the neural network
agent = Agent(net) # initialize the trainer
# start training, set total number of training steps to 20000
agent.train(total_time_steps=20000)
```

Execute python train_ppo.py in the terminal to start training. On an ordinary laptop, it takes only **a few seconds** to complete the agent's training.

TIP

OpenRL also provides command line tools that allow you to complete agent training with one command. Users only need to execute the following command in the terminal:

```
openrl --mode train --env CartPole-v1
```

Test Environment

After the agents have completed their training, we can use the `agent.act()` method to obtain actions. Just add this code snippet into your `train_ppo.py` file and visualize test results:

Main Features of OpenRL

Customizable capabilities for professionals

Configure everything via YAML

```yaml
seed: 0
lr: 7e-4
critic_lr: 7e-4
episode_length: 25
run_dir: ./run_results/
experiment_name: train_mpe
log_interval: 10
```

Use yaml

> python train_ppo.py --config mpe_ppo.yaml

```yaml
seed: 0
lr: 7e-4
critic_lr: 7e-4
episode_length: 25
run_dir: ./run_results/
experiment_name: train_mpe
log_interval: 10
```

Use yaml

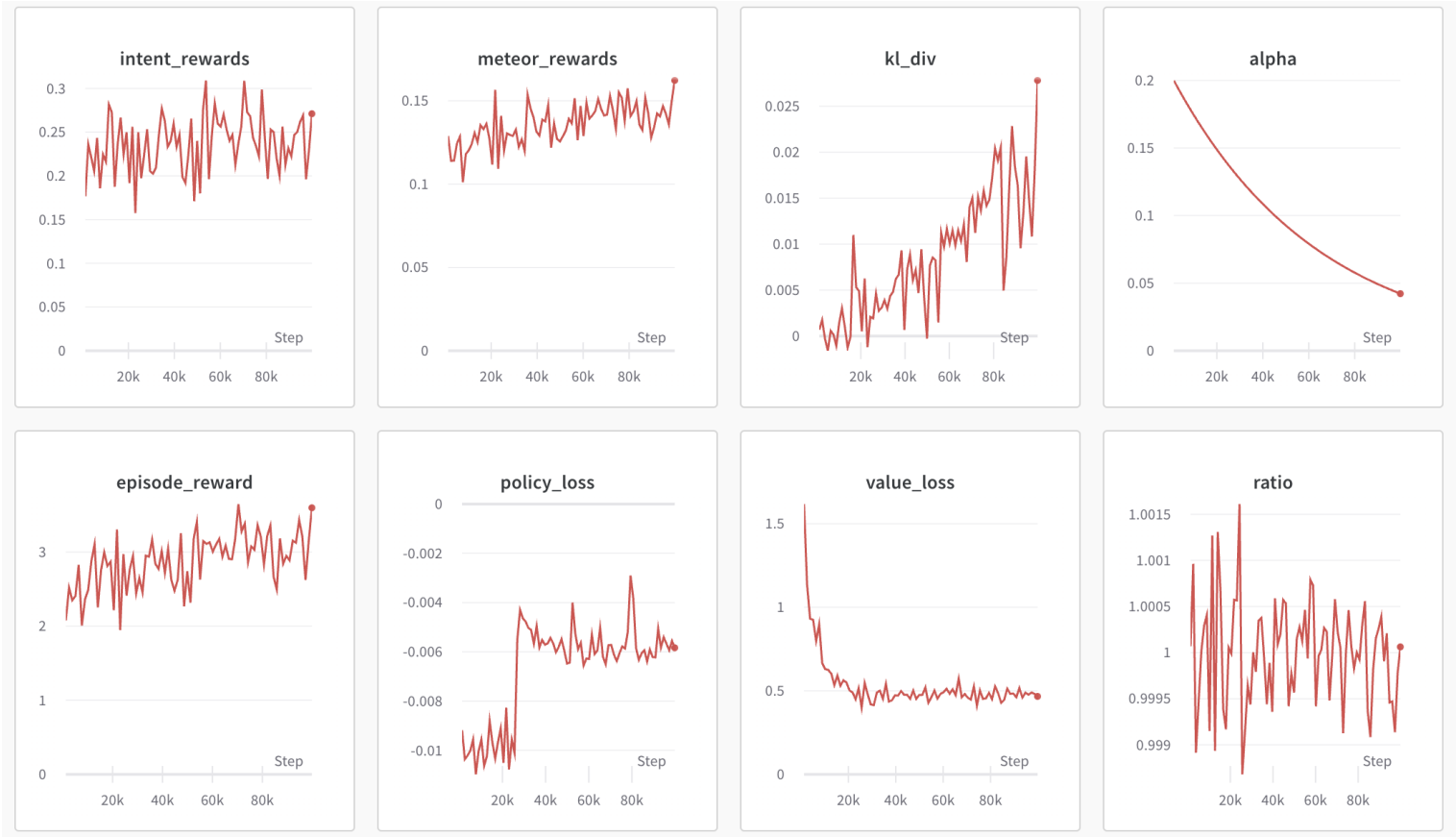
> python train_ppo.py --config mpe_ppo.yaml

> python train_ppo.py --seed 1 --lr 5e-4

Main Features of OpenRL

Customizable capabilities for professionals

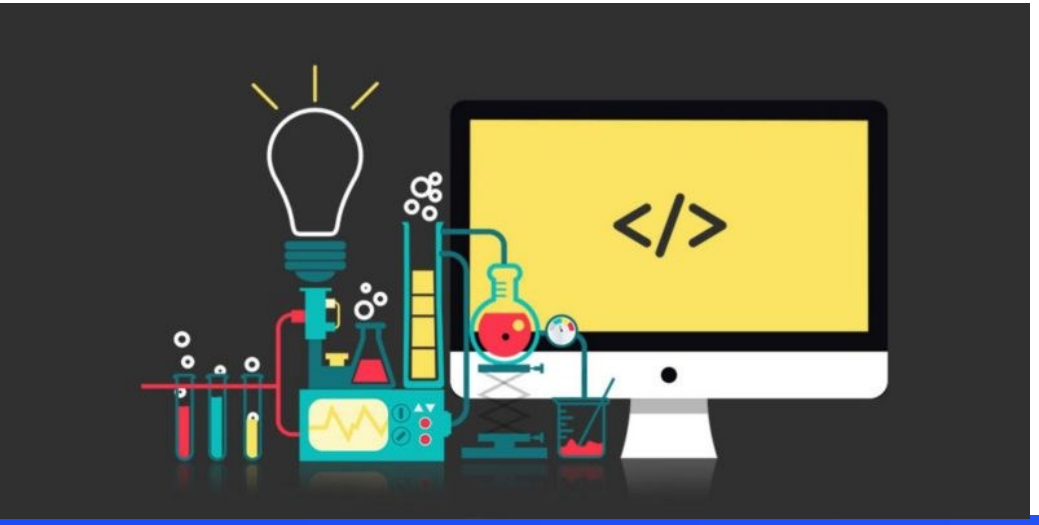
Track your experiments via Wandb

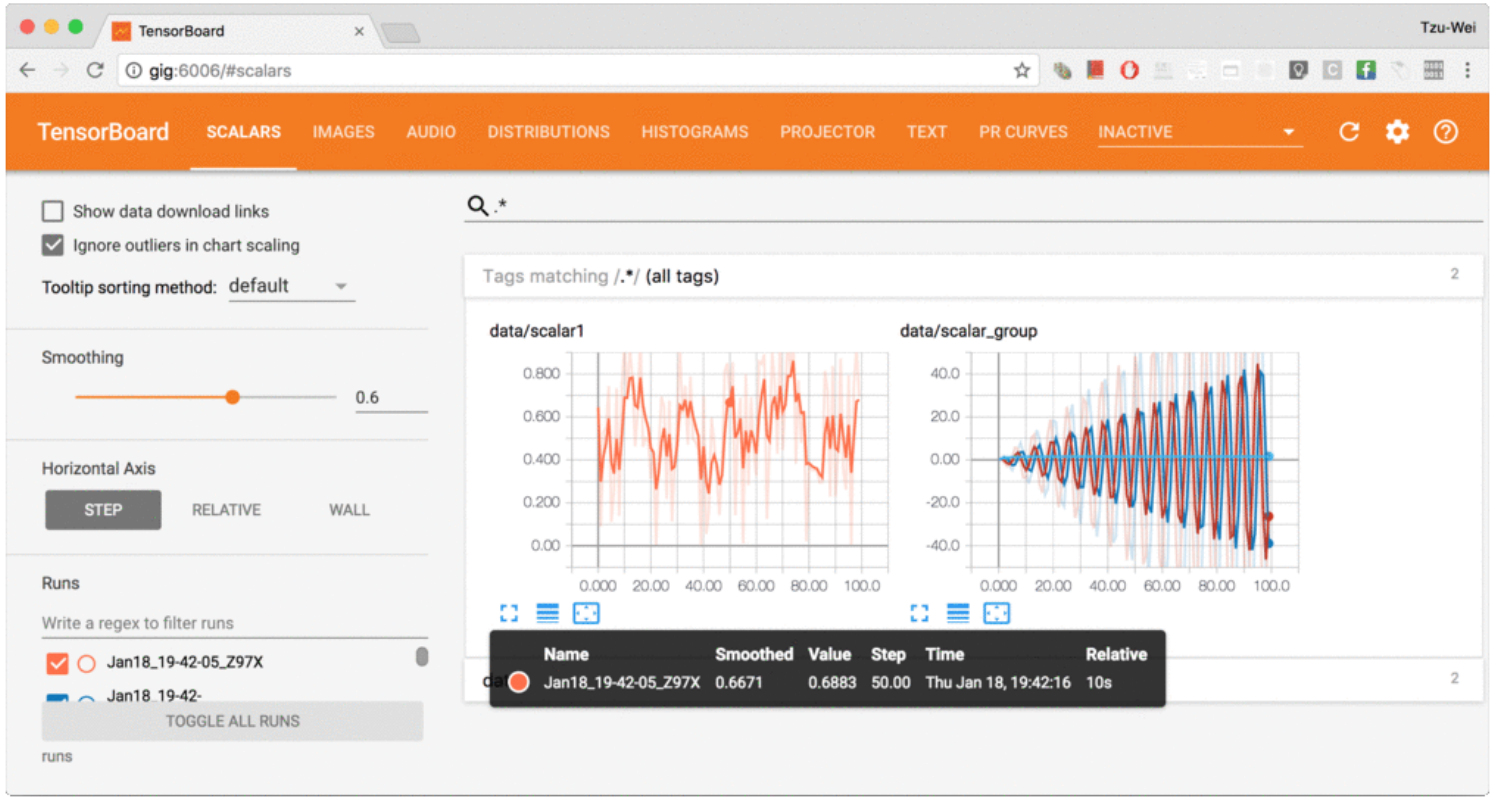


Main Features of OpenRL

Customizable capabilities for professionals

Track your experiments via Tensorboard



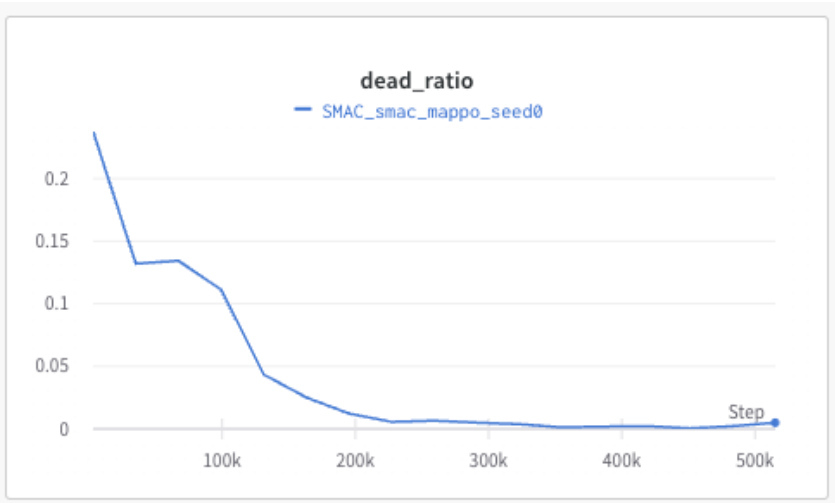


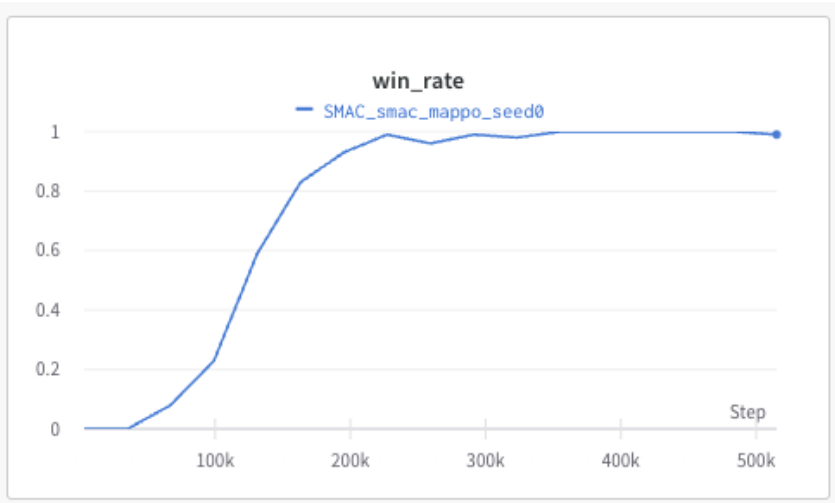
Customize Wandb Output

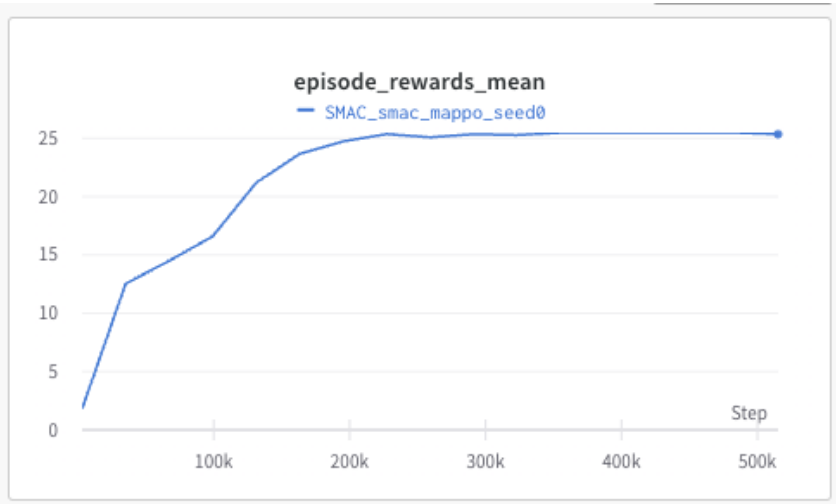
```python
class SMACInfo(EPS_RewardInfo):
    def __init__(self, *args, **kwargs):
        super().__init__(*args, **kwargs)
        self.win_history = deque(maxlen=100)


    def statistics(self, buffer: Any) -> Dict[str, Any]:
```

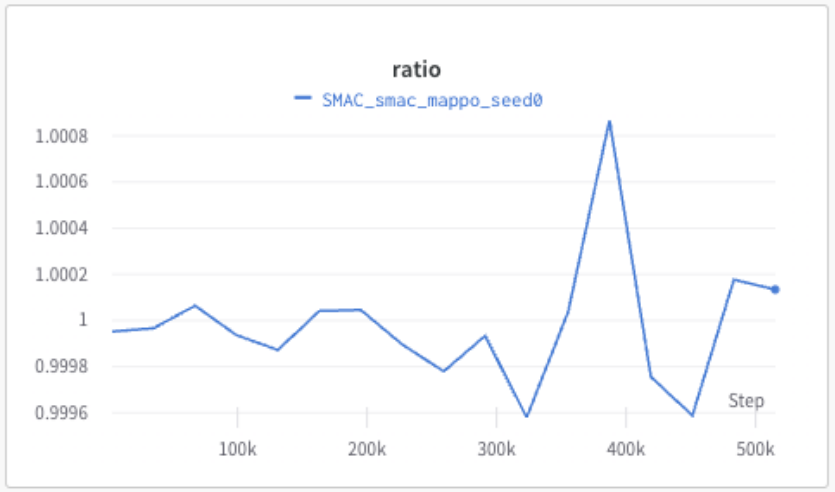
https://github.com/OpenRL-Lab/openrl/blob/main/examples/smac/custom_vecinfo.py

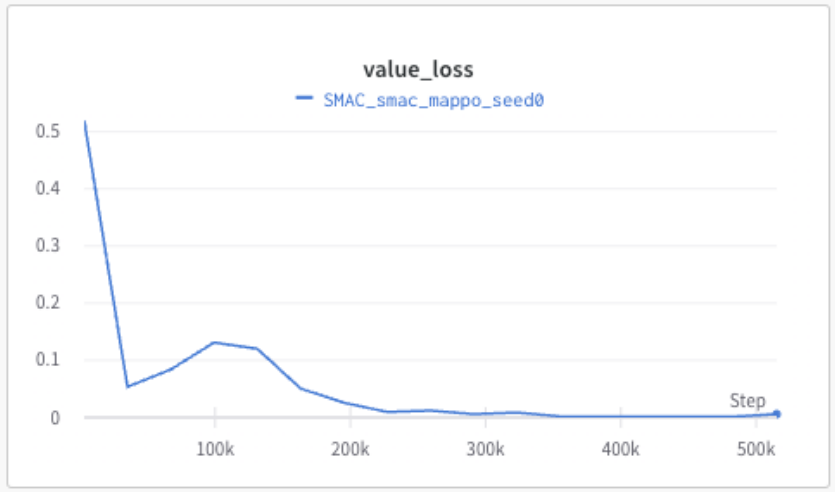
Customize Wandb Output

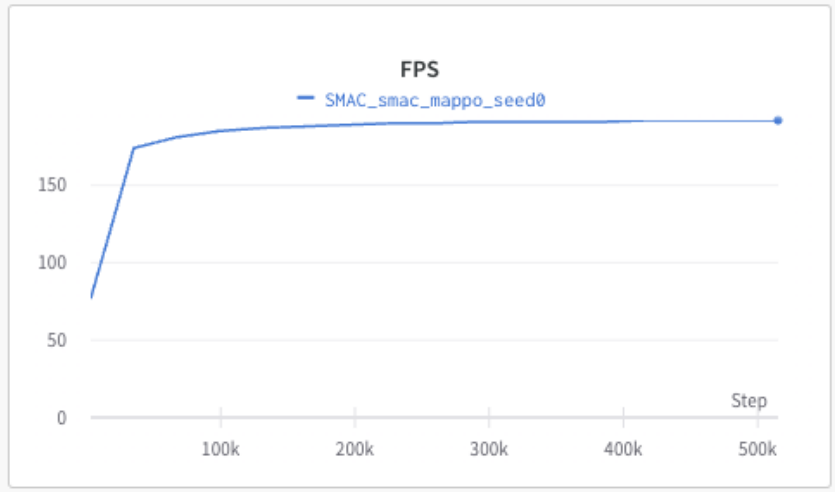








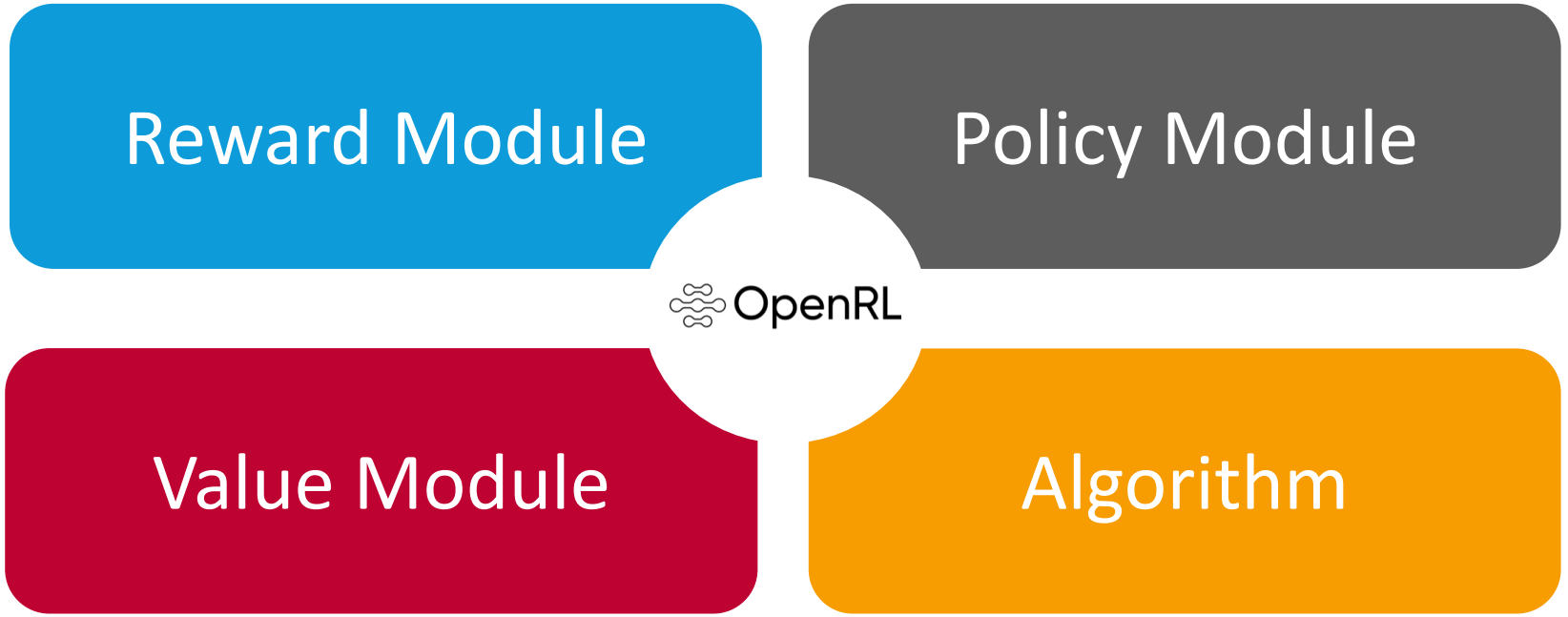




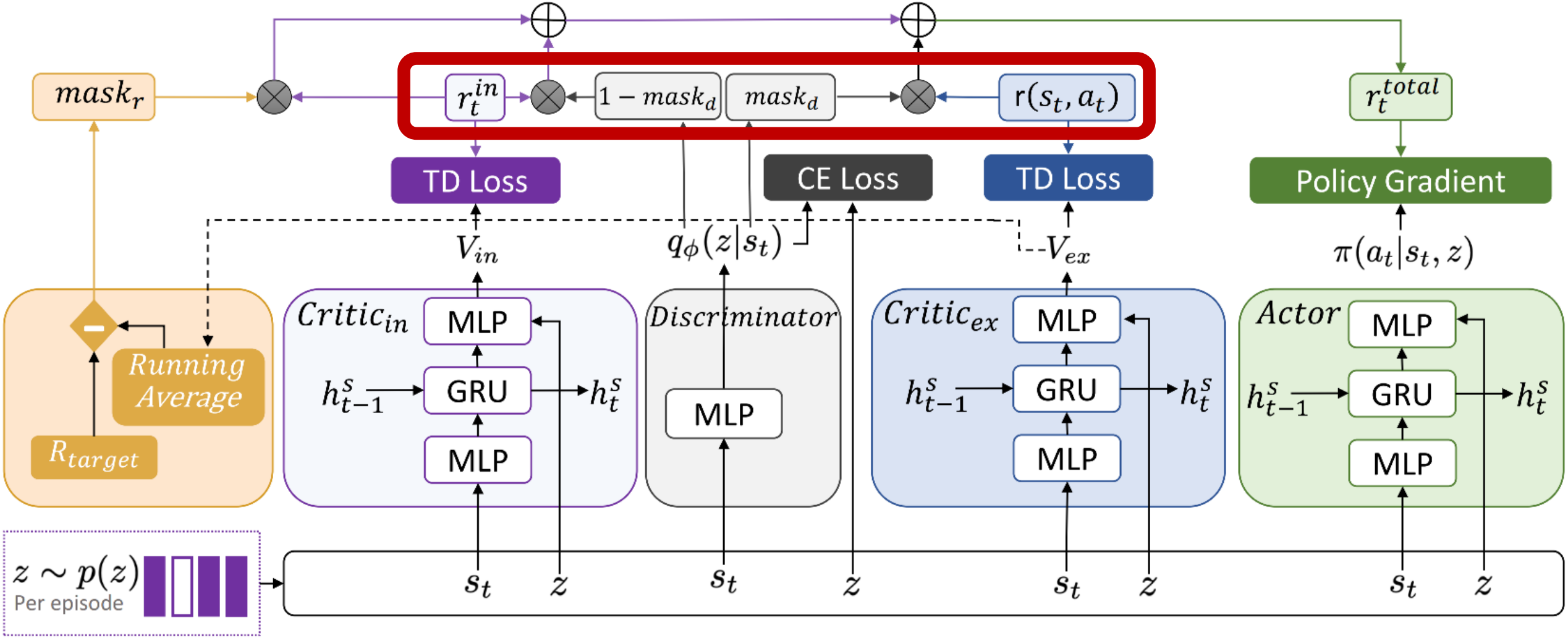
Main Features of OpenRL

Customizable capabilities for professionals

Abstract & Modularized Design

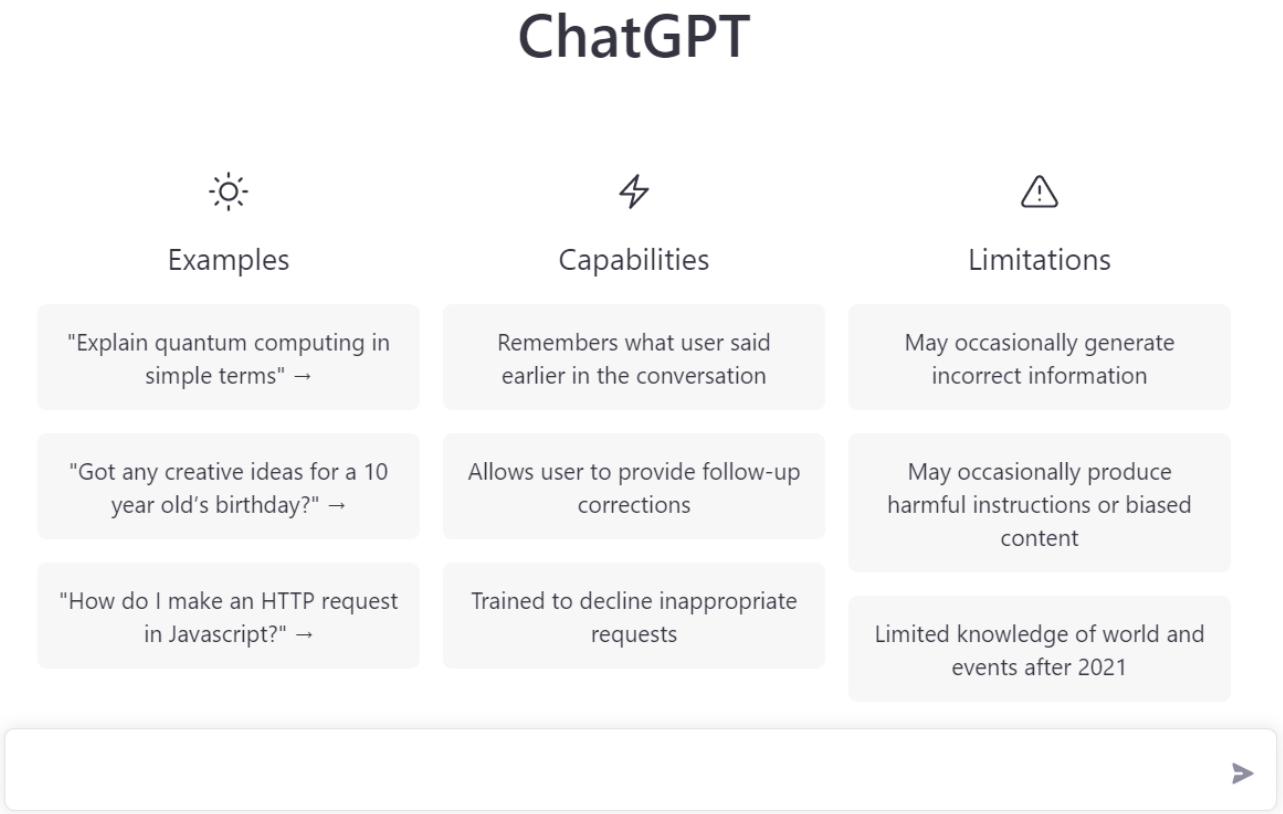


Customize Reward Model



Chen, Wenze, et al. "DGPO: Discovering Multiple Strategies with Diversity-Guided Policy Optimization." arXiv preprint arXiv:2207.05631 (2022).

Customize Reward Model



Step 3

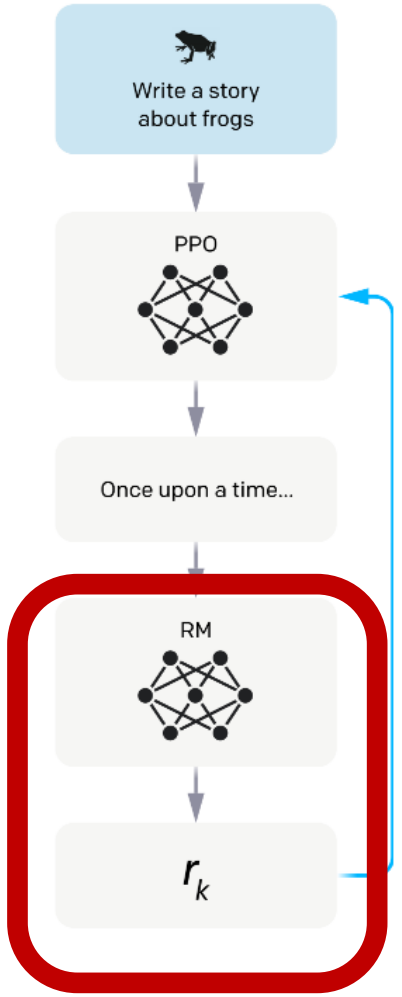
Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



Customize Reward Model

```python
class BaseReward(object):
    def __init__(self):
        self.step_reward_fn = dict()
        self.inner_reward_fn = dict()
        self.batch_reward_fn = dict()
```

Customize Reward Model

```
class NLPReward(BaseReward):
```

- **Intent Reward**： When the generated text by the agent is close to the expected intent, the agent can receive higher rewards.

- **METEOR Metric Reward**： METEOR is a metric used to evaluate text generation quality and can be used to measure how similar generated texts are compared with expected ones. We use this metric as feedback for rewards given to agents in order to optimize their text generation performance.

- **KL Divergence Reward**： This reward is used to limit how much text generated by agents deviates from pre-trained models and prevent issues of reward hacking.

Customize Reward Model

```
class NLPReward(BaseReward):
```

- **Intent Reward**： When the generated text by the agent is close to the expected intent, the agent can receive higher rewards.

```
self.batch_rew_funcs = {
    "intent_acc": Intent(**intent_config),
}
```

Customize Reward Model

```
class NLPReward(BaseReward):
```

- **METEOR Metric Reward**： METEOR is a metric used to evaluate text generation quality and can be used to measure how similar generated texts are compared with expected ones. We use this metric as feedback for rewards given to agents in order to optimize their text generation performance.

```
self.inner_reward_fn = {
    "meteor": Meteor(**meteor_config),
}
```

Customize Reward Model

```
class NLPReward(BaseReward):
```

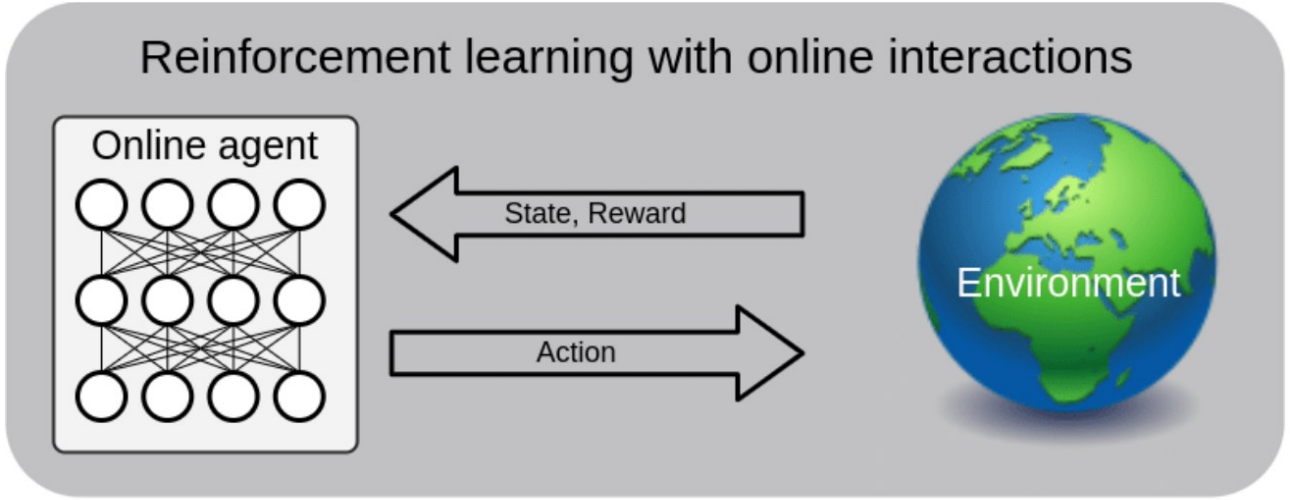
- **KL Divergence Reward**： This reward is used to limit how much text generated by agents deviates from pre-trained models and prevent issues of reward hacking.

```
self.step_rew_funcs = {
    "kl_pen": KLPenalty(**kl_config),
}
```

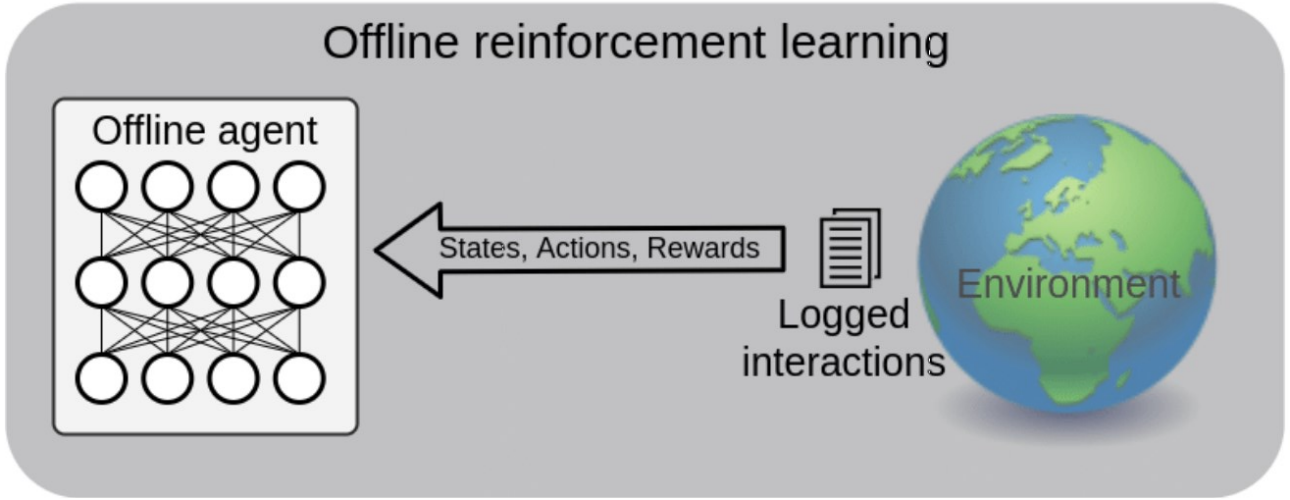
Main Features of OpenRL

Support Offline RL

Learn from Iteraction



Learn from Expert Data



Main Features of OpenRL

Support Offline RL

```
# create environment, set environment parallelism to 9
env = make("OfflineEnv", env_num=10, cfg=cfg)
# create the neural network
net = Net(env,cfg=cfg)
# initialize the trainer
agent = Agent(net)
# start training, set total number of training steps to 100000
agent.train(total_time_steps=100000)
env.close()
```

Main Features of OpenRL

Find out more

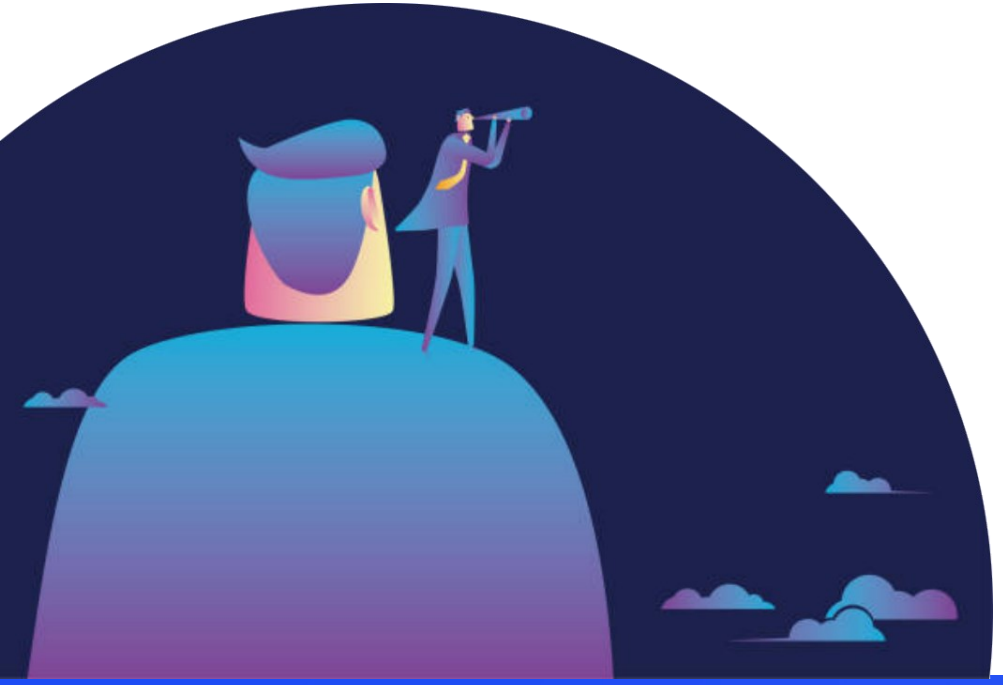
Customizable capabilities for professionals

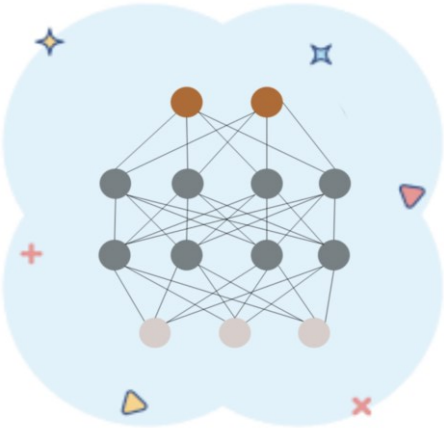
- **Dictionary observation** space support
- **Serial** or **parallel** environment training
- Support for models such as **LSTM**, **GRU**, **Transformer** etc.
- Automatic mixed precision (**AMP**) training
- Data collecting wth **half precision policy** network

Main Features of OpenRL

Build on top of others

Hugging Face



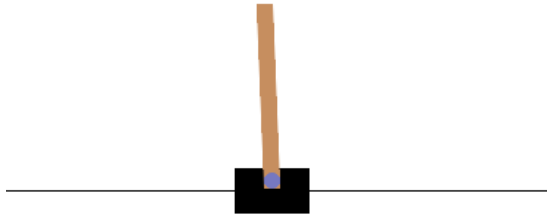


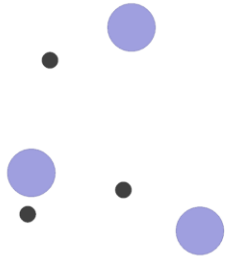
Models

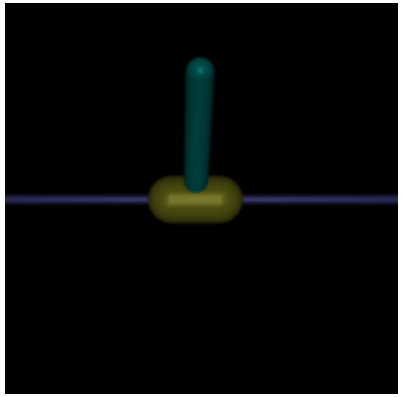
Datasets

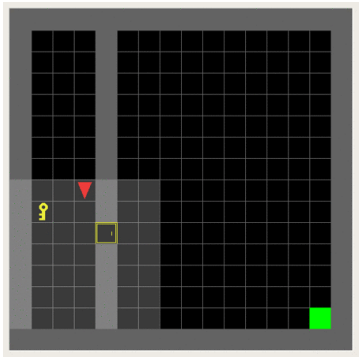
Main Features of OpenRL

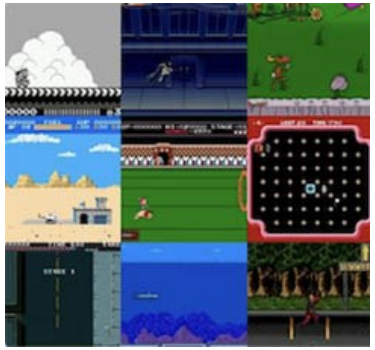
Gallery

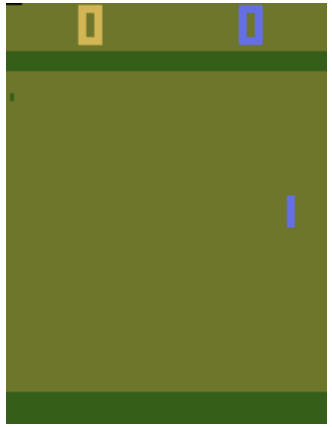












Main Features of OpenRL

High performance

Training CartPole on a laptop only takes **a few seconds**.
+17% speedup for language model training.

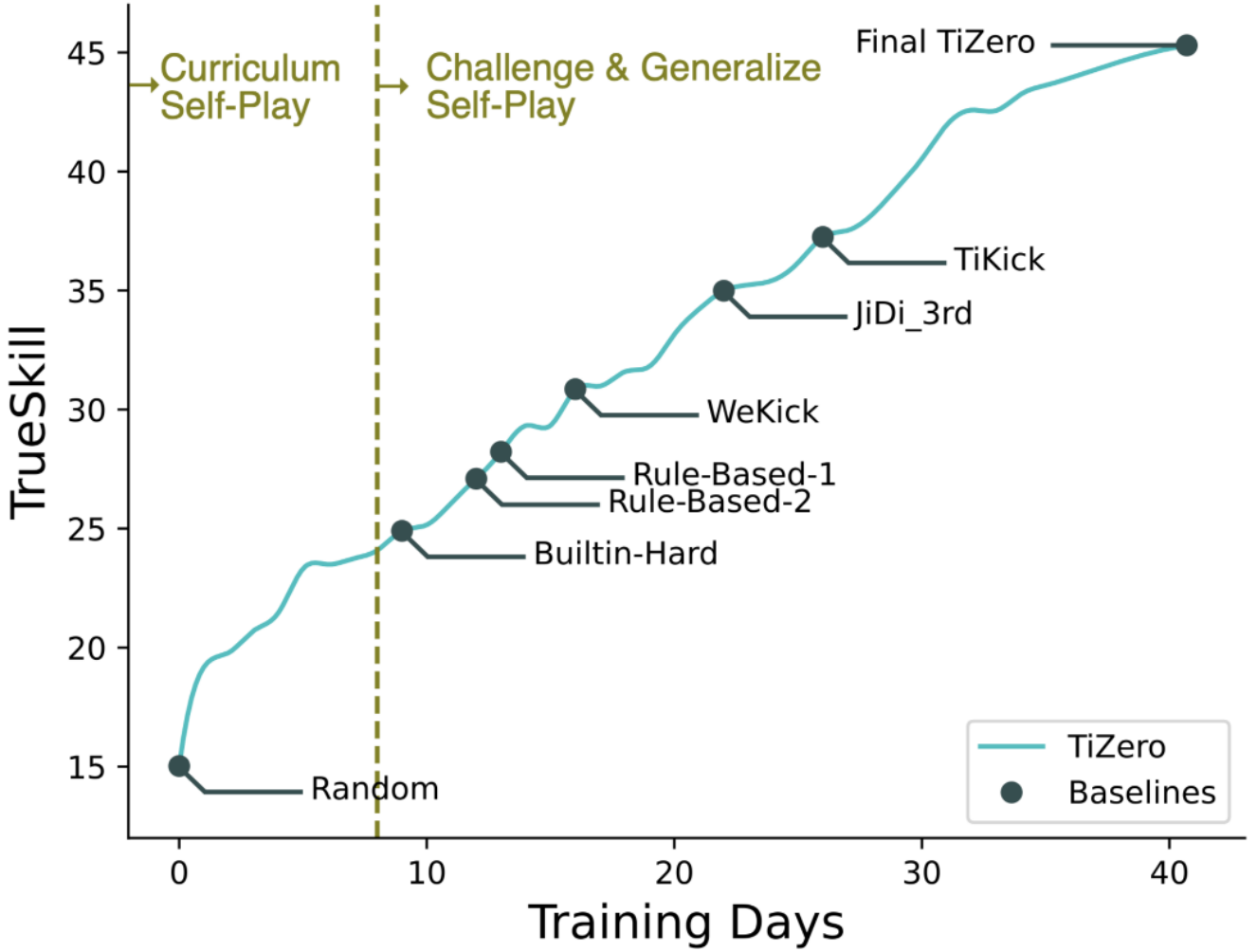
Ranking 1st on Google Research Football.
Achieving **+43%** performance improvement on LLM.

Compared with RL4LMs

	FPS(Speed)	Rouge-1	Rouge-Lsum	Meteor	SacreBLEU
Supervised Learning	None	0.164	0.137	0.234	0.063
RL4LMs	11.26	0.169	0.144	0.198	0.071
OpenRL	**13.20(+17%)**	**0.181(+10%)**	**0.153(+12%)**	**0.292(+25%)**	**0.090(+43%)**

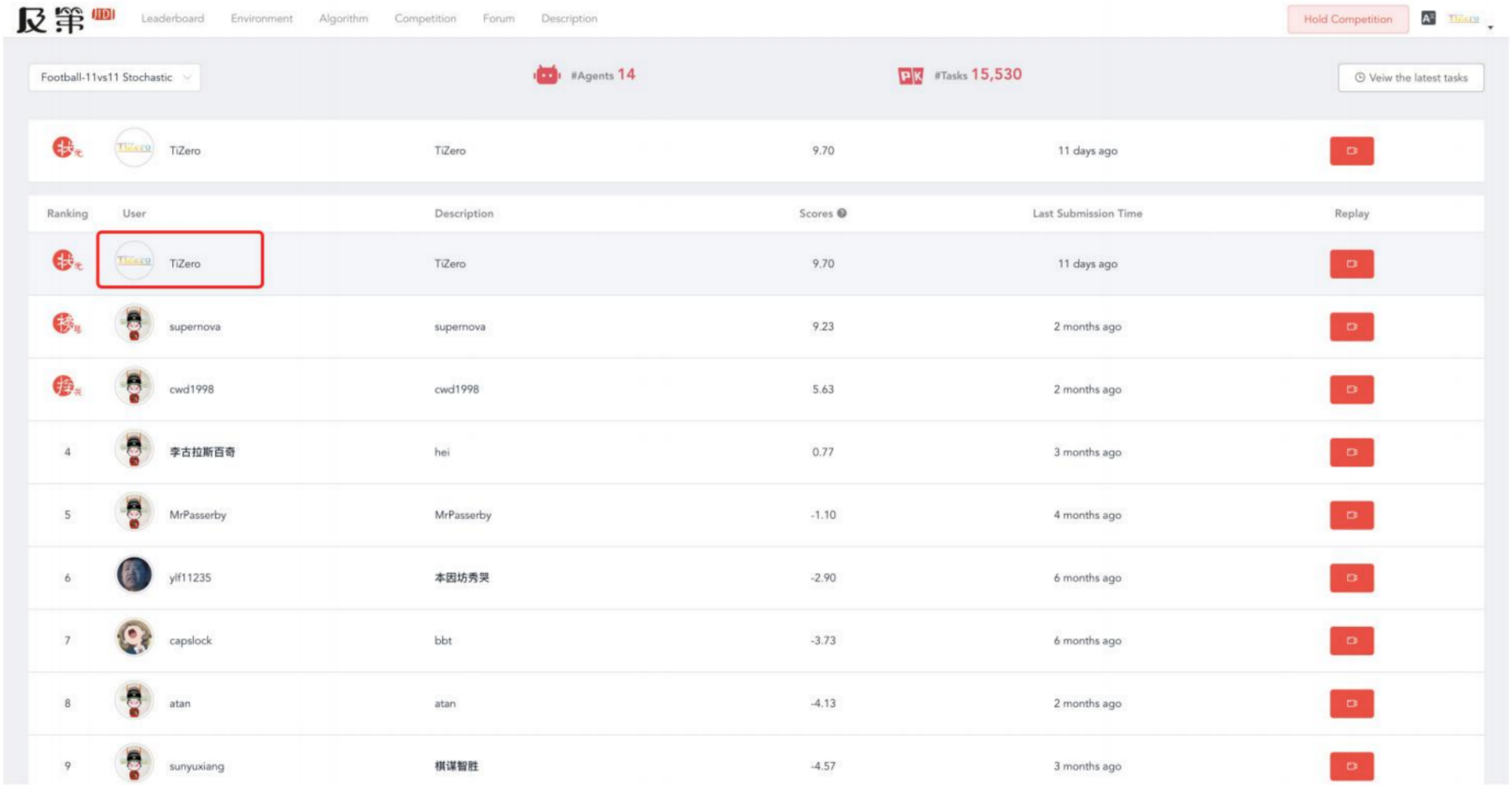
TiZero

Lin, Fanqi, et al. "TiZero: Mastering Multi-Agent Football with Curriculum Learning and Self-Play." arXiv preprint arXiv:2302.07515 (2023).



TiZero

Lin, Fanqi, et al. "TiZero: Mastering Multi-Agent Football with Curriculum Learning and Self-Play." arXiv preprint arXiv:2302.07515 (2023).



TiZero

Lin, Fanqi, et al. "TiZero: Mastering Multi-Agent Football with Curriculum Learning and Self-Play." arXiv preprint arXiv:2302.07515 (2023).

PART 03
Future Release

Large-Scale RL

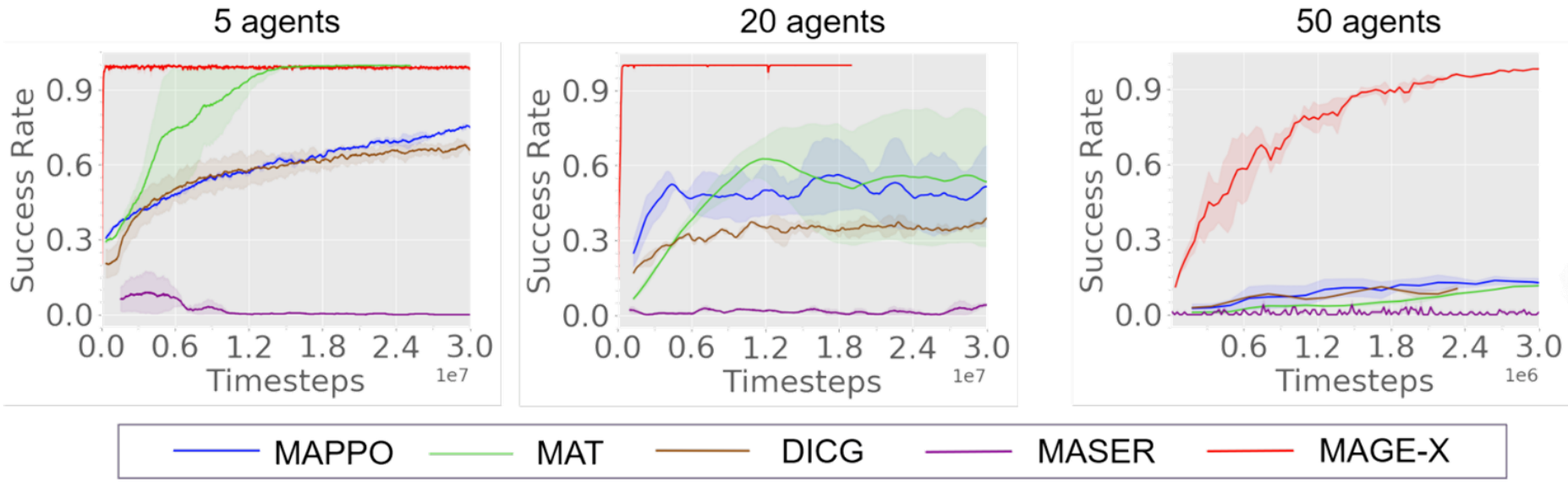
Large Model

Large Cluster

Large Population

Large-Scale RL

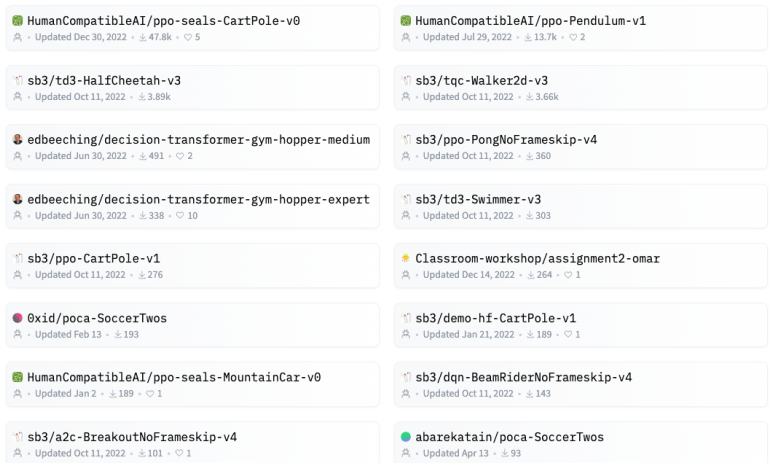
Large Population



Yang, Xinyi, et al. "Learning Graph-Enhanced Commander-Executor for Multi-Agent Navigation." arXiv preprint arXiv:2302.04094 (2023).

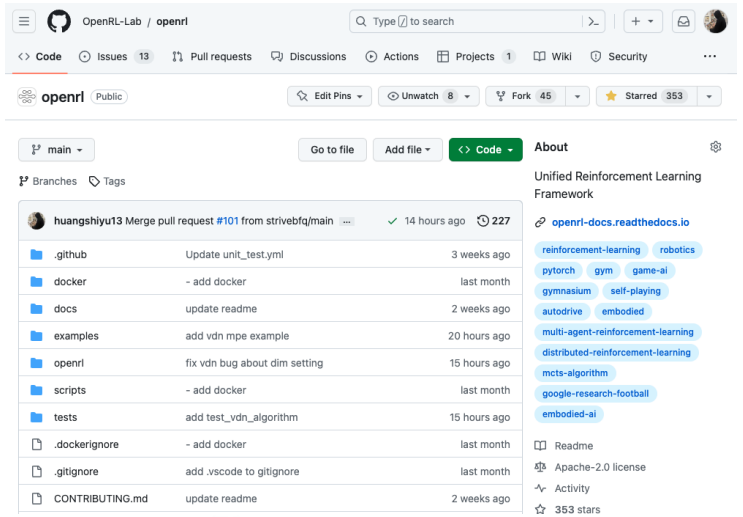
Open RL via Sharing

Share Models



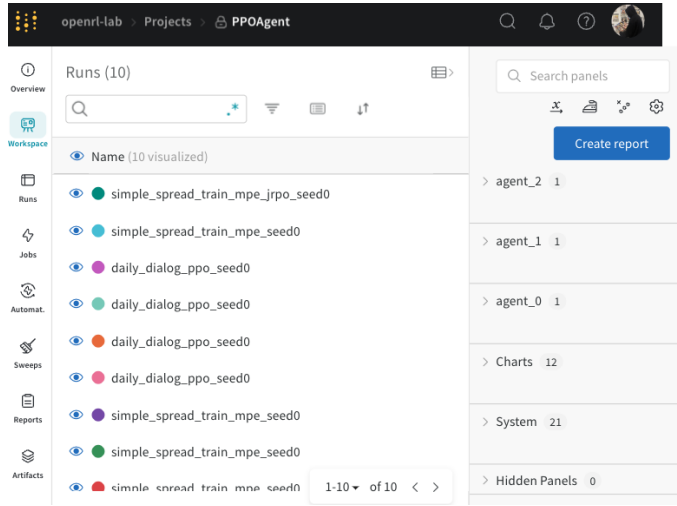
Hugging Face

Share Codes



GitHub

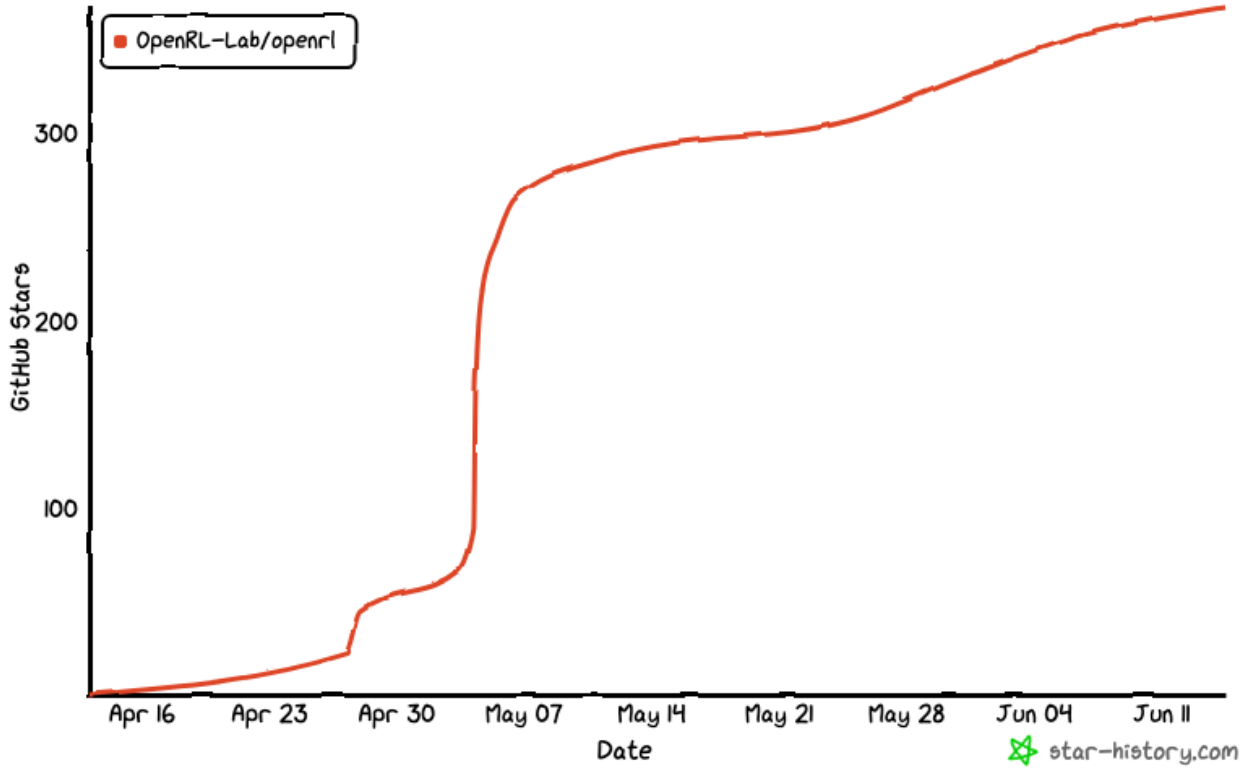
Share Results



Weights & Biases

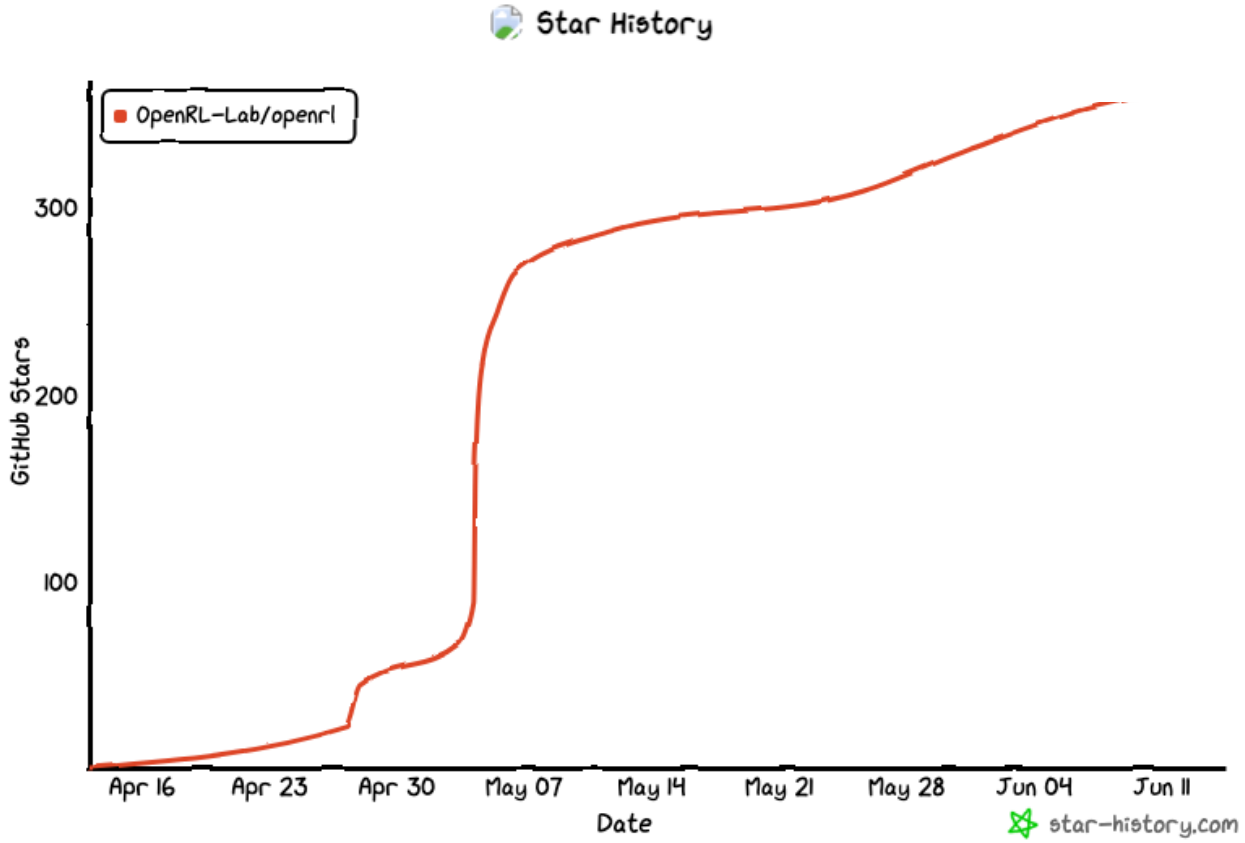
Scan the QR code to try OpenRL!

OpenRL



Visit: github.com/OpenRL-Lab/openrl

Scan the QR code to try OpenRL!



Visit: github.com/OpenRL-Lab/openrl

PART 03
OpenPlugin: Plugins for LLM

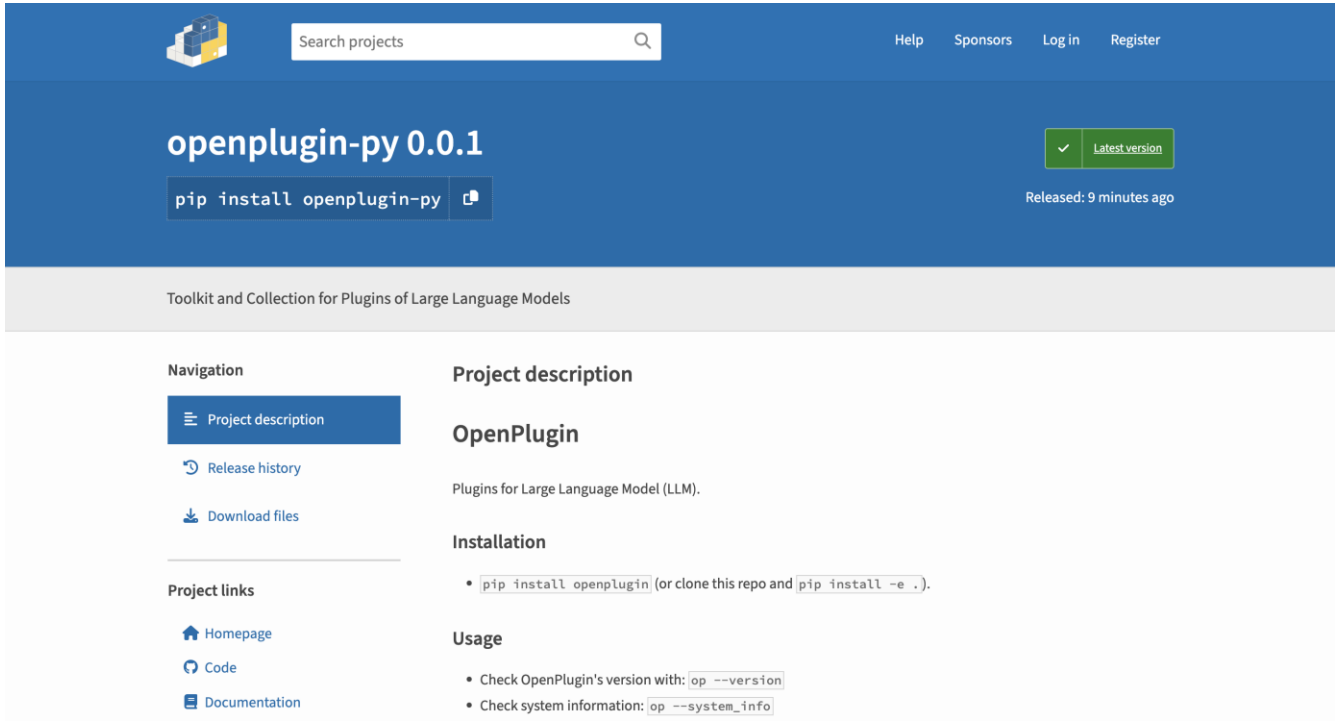
Why?

- Think about **pip** for Python package (apt/yum/brew/dnf/npm/)!
- Think about App Store.
- Standardize plugin.
- **Provide a simple way to use, share LLM plugins.**

Main Features of OpenPlugin

Installation

pip install openplugin-py



Main Features of OpenPlugin

Find out more

Usage

- install plugin: **op install** <plugin_name>
  - install locally **op install ./**
  - reinstall **op reinstall** <plugin_name>
- uninstall plugin: **op uninstall** <plugin_name>
- start to run plugin: **op run** <plugin_name>
- list installed plugins: **op list**

op is all you need!

Main Features of OpenPlugin

Find out more

Usage

- Provide config API for SageGPT/ChatGPT platform
  - can get json file via: **server_host/ai-plugin.json**
  - can get YAML file via: **server_host/openapi.yaml**

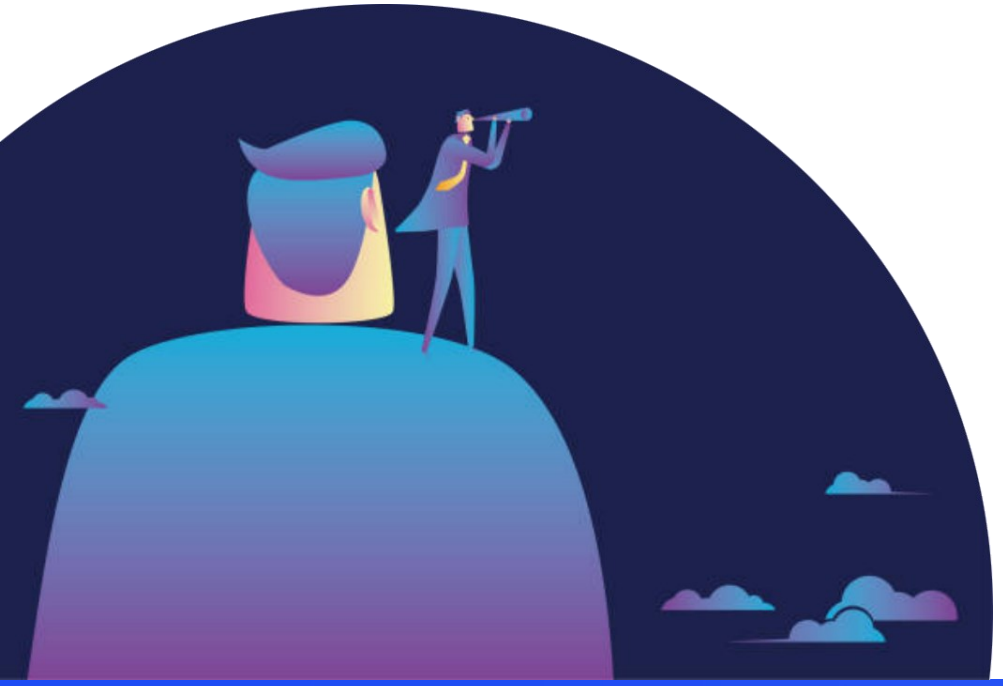
```
{
    "schema_version": "v1",
    "name_for_human": "二维码生成",
    "name_for_model": "QR code",
    "description_for_human": "这个插件可以为你生成一张二维码图片",
    "description_for_model": "这个插件可以为用户生成一张二维码图片",
    "auth": {
        "type": "none"
    },
    "api": {
        "type": "openapi",
        "url": "paint-plugin-openapi.yaml",
        "is_user_authenticated": false
    },
    "logo_url": "http://imageOcrSummary.4pd.io/logo.png",
    "contact_email": "huangsy1314@163.com",
    "legal_info_url": "http://imageOcrSummary.4pd.io/legal"
}
```

```
openapi: 3.0.0
info:
  title: QR code API
  description: 这是一个用于获取二维码图片的API。
  version: 1.0.0

servers:
  - url: http://172.24.4.12:5004

paths:
  /qrcode_image:
    get:
      summary: 获取二维码图片
      operationId: getQRcode
```

Main Features of OpenPlugin

Build on top of others

https://openrl.net/plugin-store/



Home OpenPlugin

Plugin Store

> Plugins for Large Language Model.

Plugin Name	Description
ikun_plugin	I Love Kun!
todo_plugin	make todo list
QRcode_plugin	Generate QR code for you!

You can share your plugin to others!

Main Features of OpenPlugin

Plugin Store

QRcode_plugin

todo_plugin

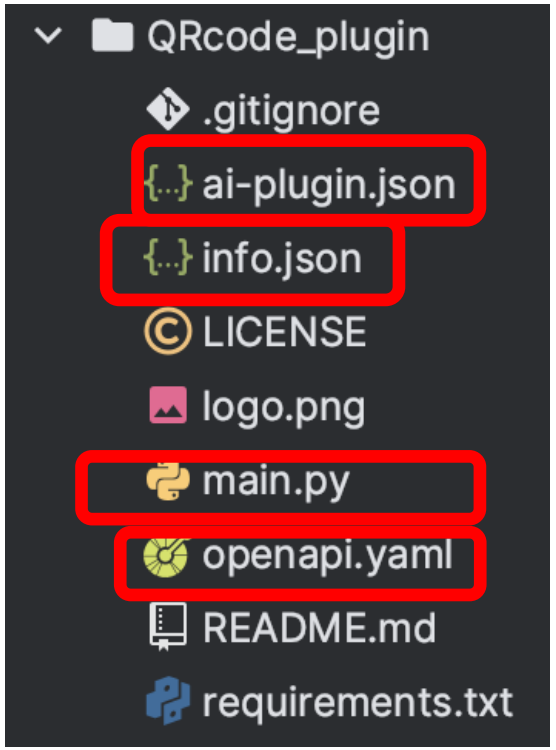
ikun_plugin

. . . .

Main Features of OpenPlugin

QRcode_plugin

Plugin Structure



Support for placeholder:

```yaml
  title: QR code API
  description: 这是一个用于获取二维码图片的API。
  version: 1.0.0

servers:
  - url: {% ROOT_URL %}

paths:
  /qrcode_image:
    get:
      summary: 获取二维码图片
      operationId: getQRcode
```

Main Features of OpenPlugin

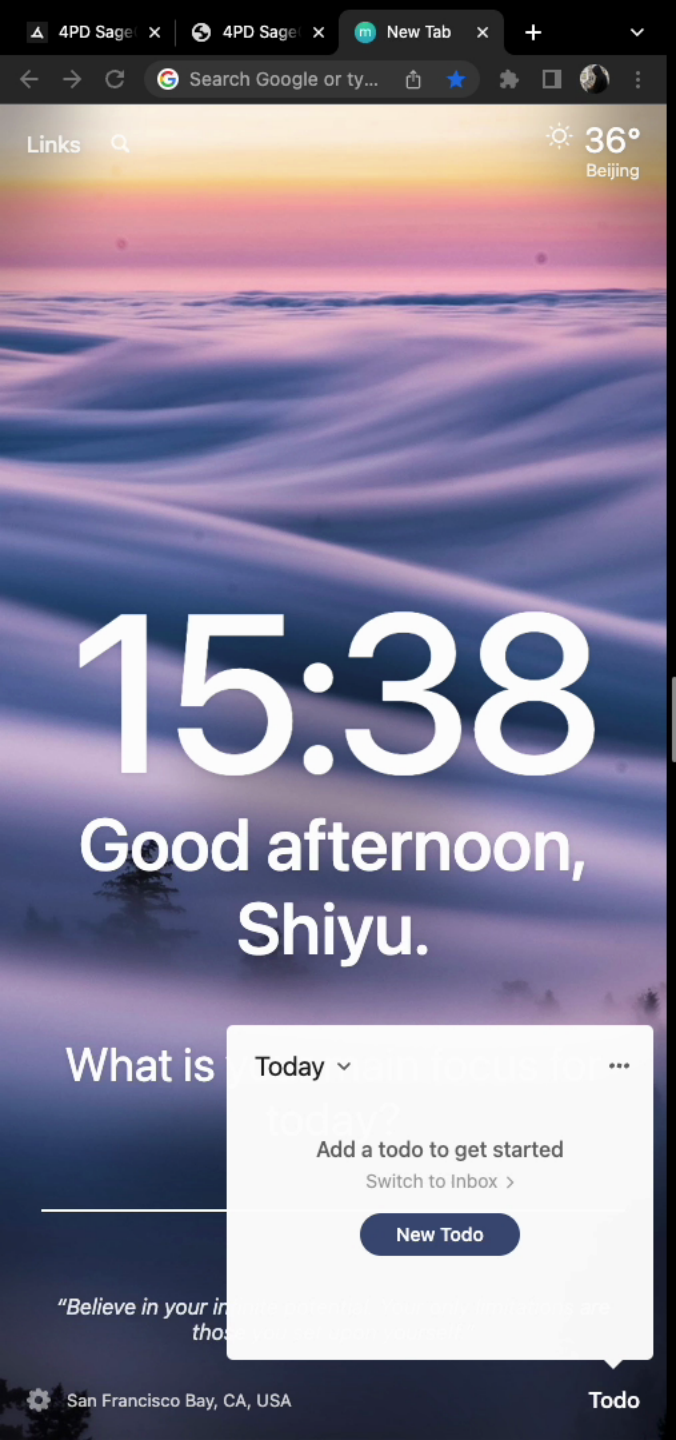
How to use QRcode_plugin

- Step 0: Find a server
- Step 1: **pip install openplugin-py**
- Step 2: **op install QRcode_plugin**
- Step 3: **op run QRcode**
- Step 4: Get the **json** and **YAML** file
- Step 5: Register plugin to SageGPT or ChatGPT website
- Step6: Finished! Have fun!

Main Features of OpenPlugin

QRcode_plugin

Demo



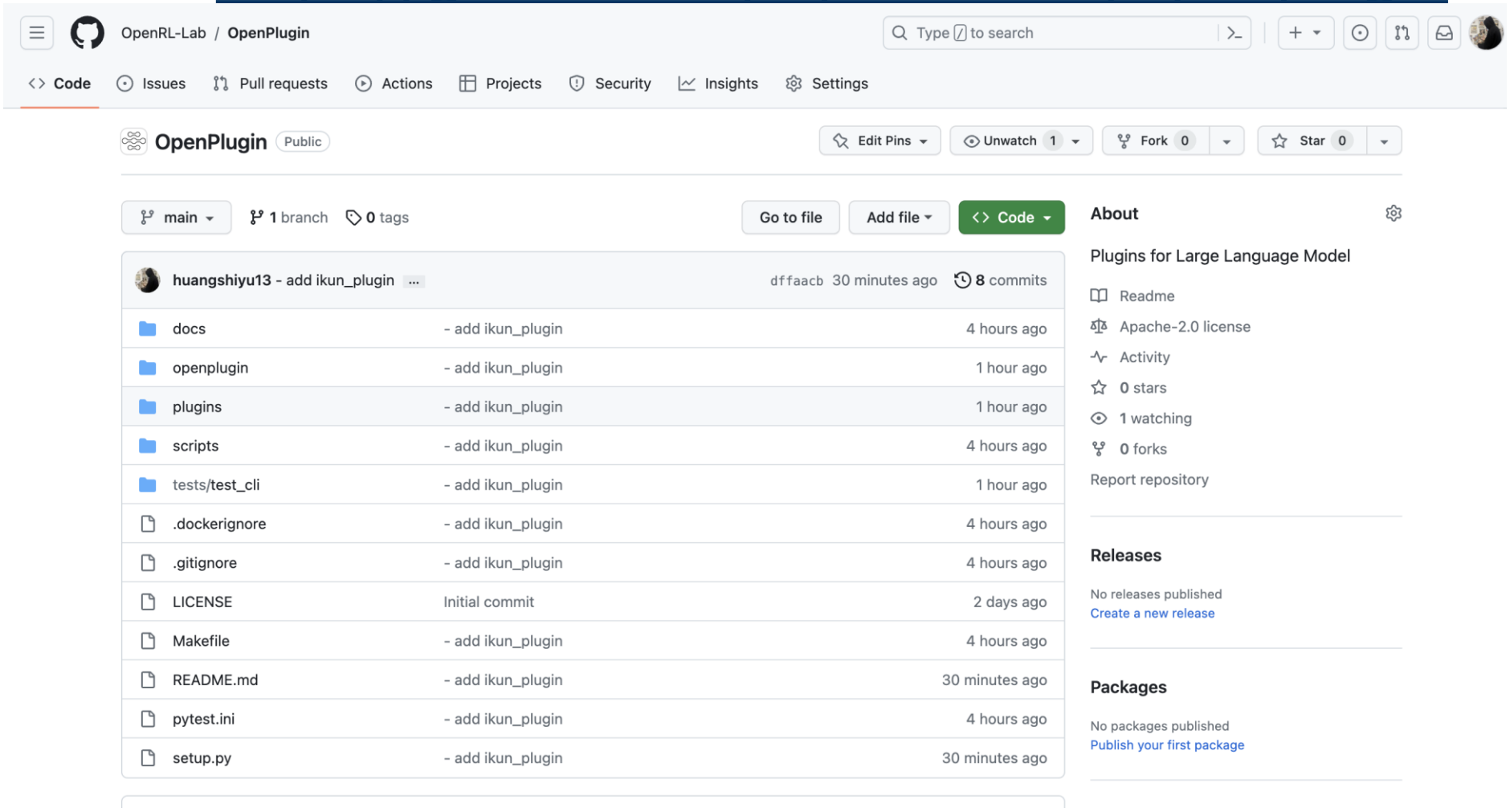
tmux (ssh) ⌘1 | data_server@pai-worker3: /data/data_server (ssh) ⌘2 | tmux (ssh) ⌘3

```
(ai3)
huangshiyu@m5-autorl-dev02:~/tmp
▶
```

[op] 0:~/tmp* "m5-autorl-dev02" 15:34 18-Jul-23

Try OpenPlugin, Click Star!

Visit: https://github.com/OpenRL-Lab/openplugin



AiDD AI+ 软件研发数字峰会
AI+ software Development Digital summit

感 谢 聆 听