Reinforcement Learning

Training Gym's LunarLander with different scenarios and models

Reinforcement Learning

FEUP-M.EIC009-2022/2023-2S

Group B
Diogo Costa - up201906731
Francisco Colino - up201905405
Henrique Sousa - up201906681

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Proposal

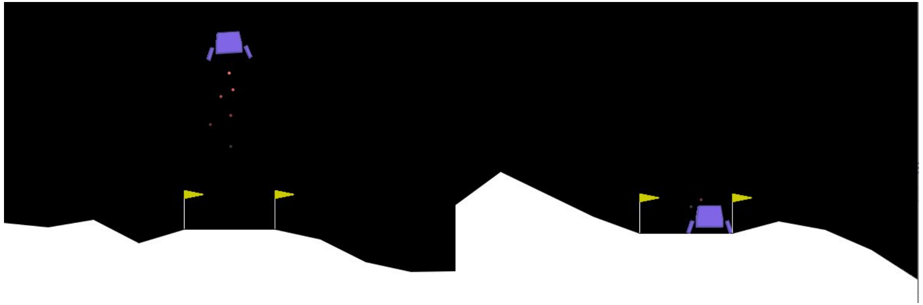
We want to explore:

- Stable baselines models and capacities.
- Gym API and environments.
- Reinforcement Learning.

Approach:

- Choose a base environment.
- Test different models.
- Test different environments.

Environment - Lunar Lander V2



- Action space -> Box(-1,1,(2,))
  - Two floats between -1 and 1
  - Main engine(fires if value > 0.5)
  - Side engines (left fires if value > 0.5, right fires if value < 0.5)
- Observation space -> (8,)
  - Lander coordinates, and velocity
  - Lander angle and angular velocity
  - If the legs contact the ground or not
- State:
  - Initial: Center of screen, with a random velocity
  - During: observation
  - Done:
    - Lander crashes
    - Lander disappears
    - Lander at rest

- Rewards
  - Negative:
    - Firing engines
    - Moving away from landing pad
    - Crashing
  - Positive
    - Moving in the direction of the landing pad
    - Finishing at rest

Algorithms applied

- A2C - Advantage Actor-Critic
- DDPG - Deep Deterministic Policy Gradient
- PPO - Proximal Policy Optimization
- SAC - Soft Actor-Critic

Training

- Per model per environment
- 15 episodes of 10k iterations
- 150k iterations

Environments used

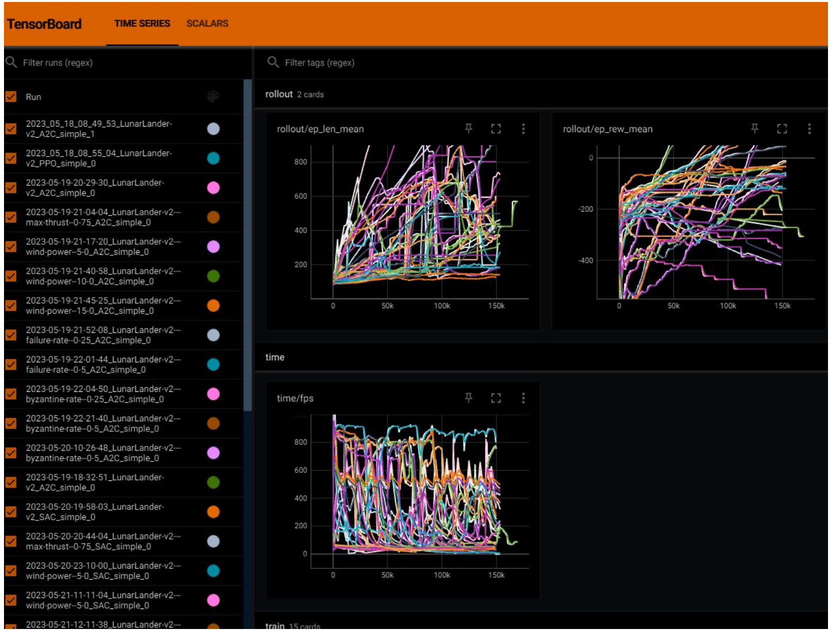
- Max thrust = 75%
- Wind power = 5
- Wind power = 10
- Wind power = 15
- Failure Rate = 5%
- Failure Rate = 10%
- Failure Rate = 25%
- Failure Rate = 50%
- Byzantine Rate = 25%
- Byzantine Rate = 50%

Visualization during and after training

- Models have a verbose option:
  - Too many prints cluttering the terminal.

Implemented our own mechanisms:

- Training by episode:
  - Variable number of episodes and iterations per episode.
  - Episode conclusion is displayed.
- Tensorboard:
  - Real time logging.
  - After the fact visualization.
- Using models:
  - Validating learning.



☑ Run

☑ 2023_05_18_08_49_53_LunarLander-v2_A2C_simple_1

☑ 2023_05_18_08_55_04_LunarLander-v2_PPO_simple_0

☑ 2023-05-19-20-29-30_LunarLander-v2_A2C_simple_0

☑ 2023-05-19-21-04-04_LunarLander-v2---max-thrust--0-75_A2C_simple_0

☑ 2023-05-19-21-17-20_LunarLander-v2---wind-power--5-0_A2C_simple_0

☑ 2023-05-19-21-40-58_LunarLander-v2---wind-power--10-0_A2C_simple_0

☑ 2023-05-19-21-45-25_LunarLander-v2---wind-power--15-0_A2C_simple_0

☑ 2023-05-19-21-52-08_LunarLander-v2---failure-rate--0-25_A2C_simple_0

☑ 2023-05-19-22-01-44_LunarLander-v2---failure-rate--0-5_A2C_simple_0

☑ 2023-05-19-22-04-50_LunarLander-v2---byzantine-rate--0-25_A2C_simple_0

☑ 2023-05-19-22-21-40_LunarLander-v2---byzantine-rate--0-5_A2C_simple_0

☑ 2023-05-20-10-26-48_LunarLander-v2---byzantine-rate--0-5_A2C_simple_0

☑ 2023-05-19-18-32-51_LunarLander-v2_A2C_simple_0

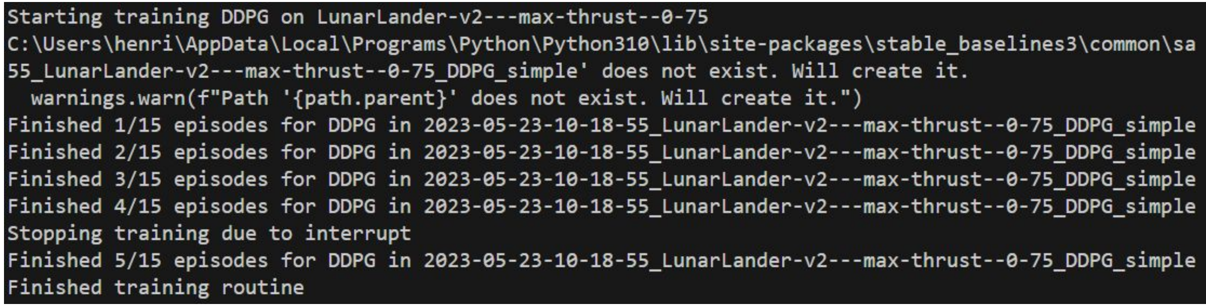
☑ 2023-05-20-19-58-03_LunarLander-v2_SAC_simple_0

☑ 2023-05-20-20-44-04_LunarLander-v2---max-thrust--0-75_SAC_simple_0

☑ 2023-05-20-23-10-00_LunarLander-v2---wind-power--5-0_SAC_simple_0

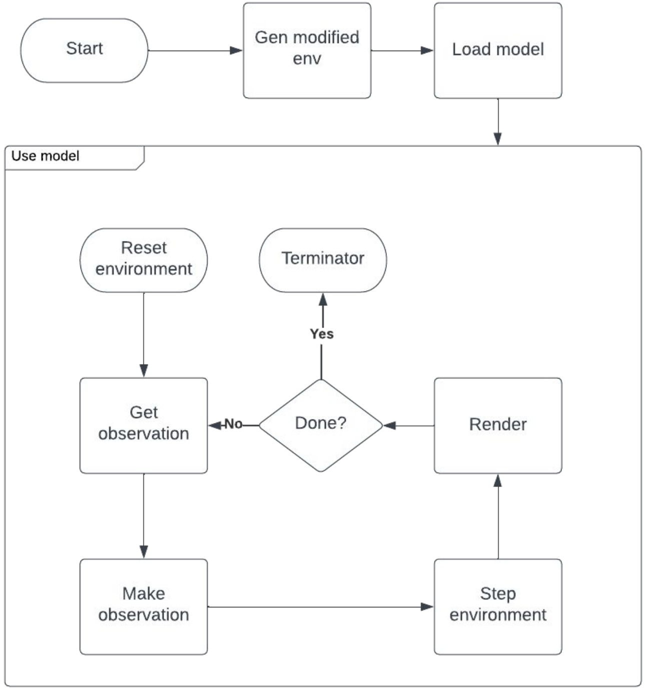
☑ 2023-05-21-11-11-04_LunarLander-v2---wind-power--5-0_SAC_simple_0

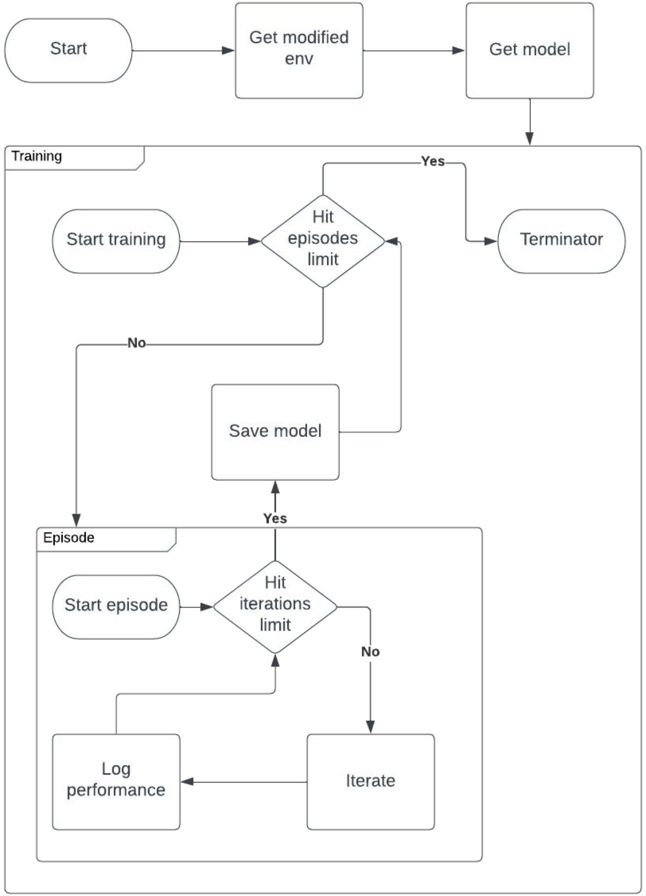
☑ 2023-05-21-12-11-38_LunarLander-v2---



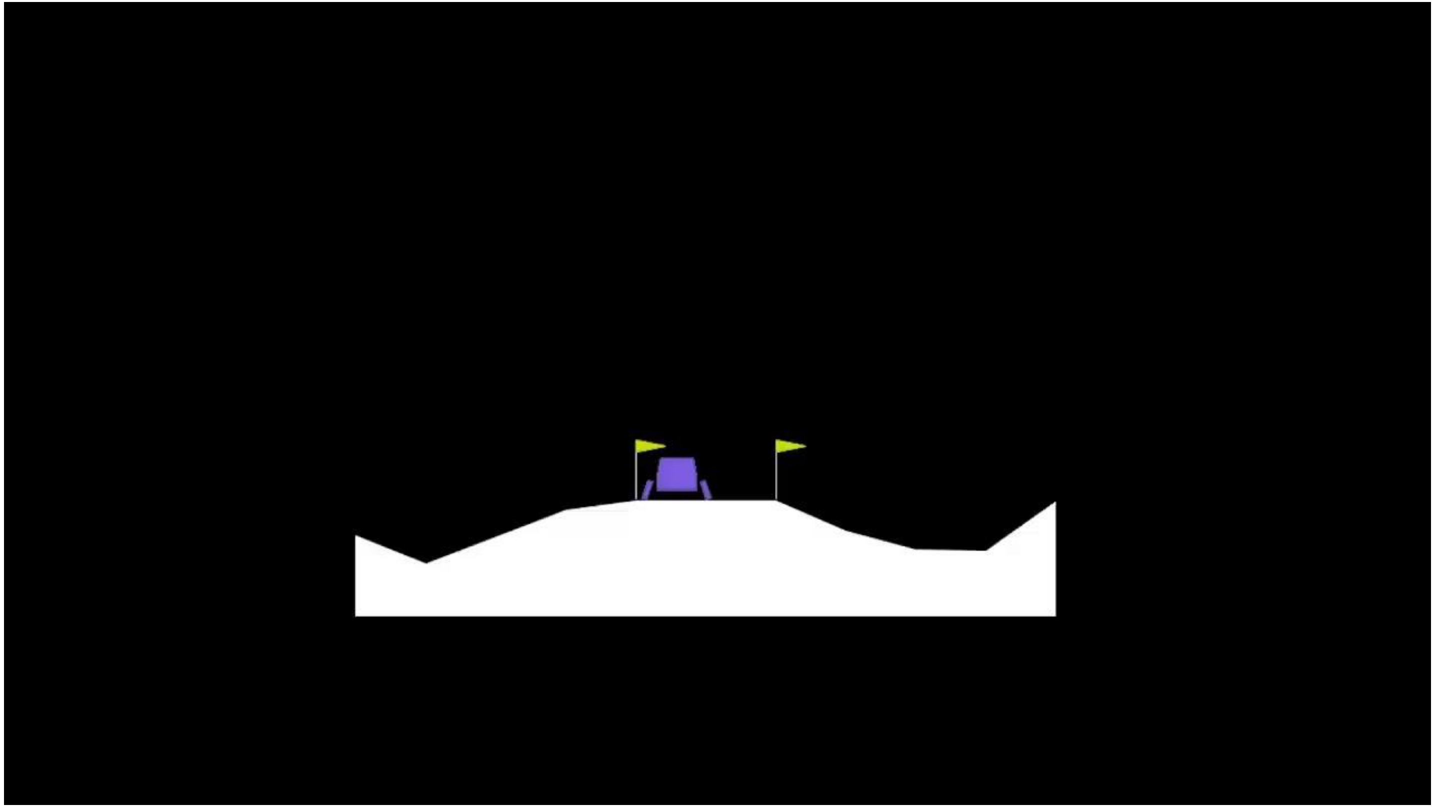
```
Starting training DDPG on LunarLander-v2---max-thrust--0-75
C:\Users\henri\AppData\Local\Programs\Python\Python310\lib\site-packages\stable_baselines3\common\sa
55_LunarLander-v2---max-thrust--0-75_DDPG_simple' does not exist. Will create it.
  warnings.warn(f"Path '{path.parent}' does not exist. Will create it.")
Finished 1/15 episodes for DDPG in 2023-05-23-10-18-55_LunarLander-v2---max-thrust--0-75_DDPG_simple
Finished 2/15 episodes for DDPG in 2023-05-23-10-18-55_LunarLander-v2---max-thrust--0-75_DDPG_simple
Finished 3/15 episodes for DDPG in 2023-05-23-10-18-55_LunarLander-v2---max-thrust--0-75_DDPG_simple
Finished 4/15 episodes for DDPG in 2023-05-23-10-18-55_LunarLander-v2---max-thrust--0-75_DDPG_simple
Stopping training due to interrupt
Finished 5/15 episodes for DDPG in 2023-05-23-10-18-55_LunarLander-v2---max-thrust--0-75_DDPG_simple
Finished training routine
```

Pipeline


Start
Gen modified env
Load model
Use model
Reset environment
Terminator
Yes
Get observation
No
Done?
Render
Make observation
Step environment


Start
Get modified env
Get model
Training
Yes
Start training
Hit episodes limit
Terminator
No
Save model
Yes
Episode
Start episode
Hit iterations limit
No
Log performance
Iterate

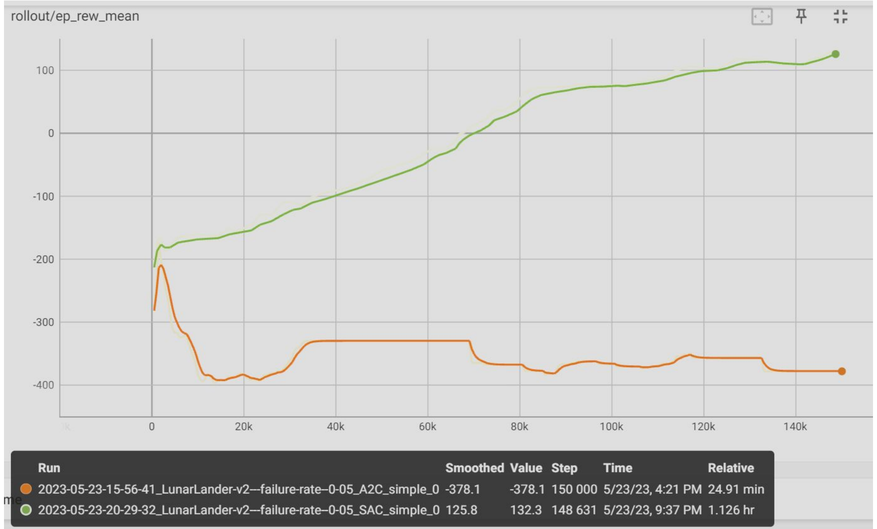
Demo



SAC with wind power 15.0

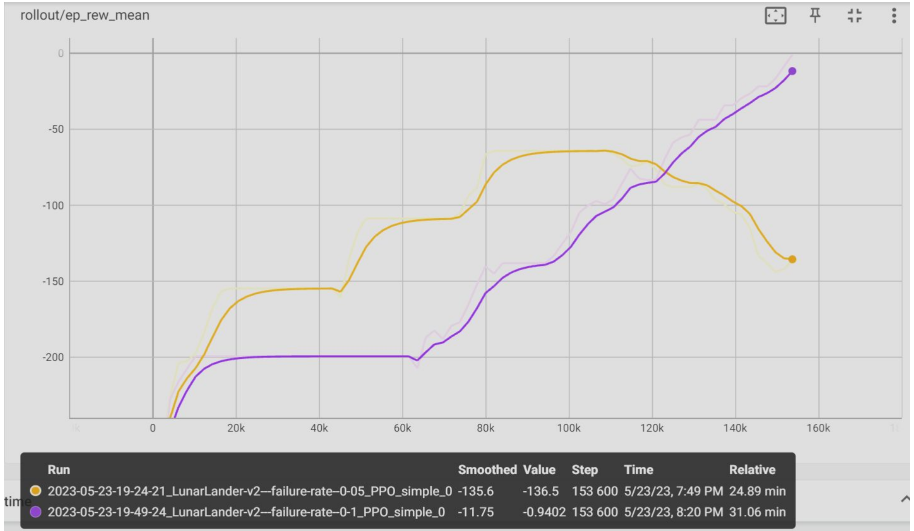
Analysis

- Comparing some models considering a failure rate in the environment:



	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-23-15-56-41_LunarLander-v2---failure-rate--0-05_A2C_simple_0	-378.1	-378.1	150 000	5/23/23, 4:21 PM	24.91 min
●	2023-05-23-20-29-32_LunarLander-v2---failure-rate--0-05_SAC_simple_0	125.8	132.3	148 631	5/23/23, 9:37 PM	1.126 hr

Orange: A2C Failure rate 0.05
Green: SAC Failure rate 0.05

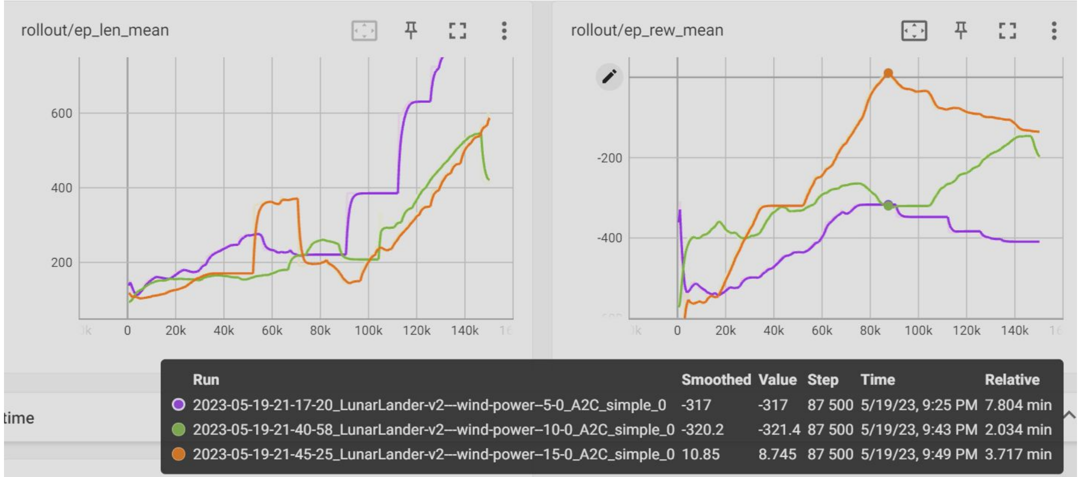


	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-23-19-24-21_LunarLander-v2---failure-rate--0-05_PPO_simple_0	-135.6	-136.5	153 600	5/23/23, 7:49 PM	24.89 min
●	2023-05-23-19-49-24_LunarLander-v2---failure-rate--0-1_PPO_simple_0	-11.75	-0.9402	153 600	5/23/23, 8:20 PM	31.06 min

Orange: PPO Failure rate 0.05
Purple: PPO Failure rate 0.1

Analysis

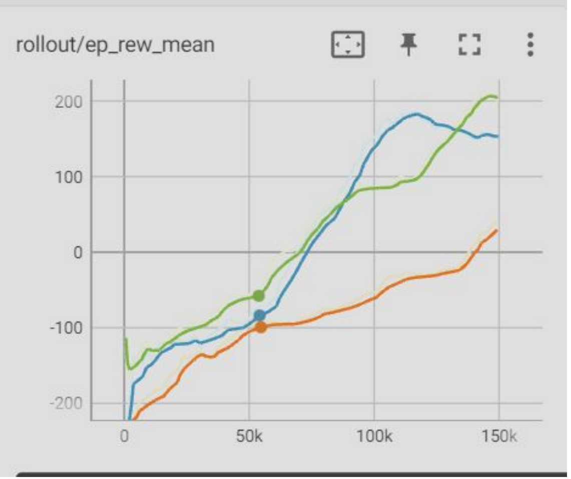
- Comparing episode length



	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-19-21-17-20_LunarLander-v2---wind-power--5-0_A2C_simple_0	-317	-317	87 500	5/19/23, 9:25 PM	7.804 min
●	2023-05-19-21-40-58_LunarLander-v2---wind-power--10-0_A2C_simple_0	-320.2	-321.4	87 500	5/19/23, 9:43 PM	2.034 min
●	2023-05-19-21-45-25_LunarLander-v2---wind-power--15-0_A2C_simple_0	10.85	8.745	87 500	5/19/23, 9:49 PM	3.717 min

Purple: A2C Wind power 5.0
Green: A2C Wind power 10.0
Orange: A2C Wind power 15.0

- Comparing models



3 version of SAC

Conclusion and future work

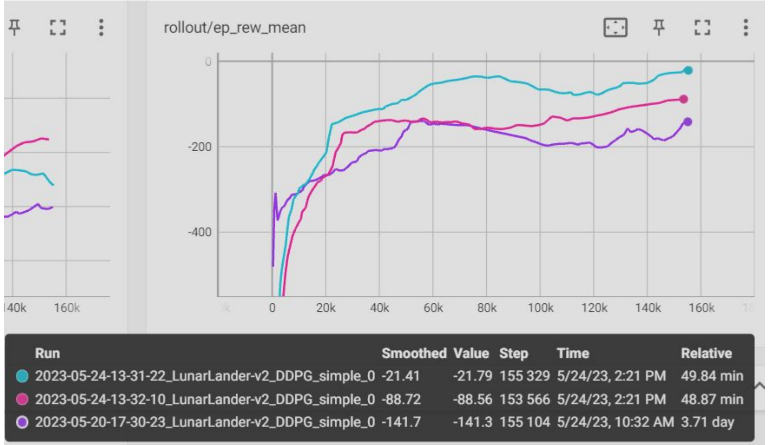
- There are libraries that make it incredibly easy to experiment with Reinforcement Learning:
    - It is even possible to get decent to good results on simple tasks.
- An adequate methodology can lead to better results and faster development.
- Training models can take a long time.
- Not all models are created equal:
    - There is enough randomization that the same conditions may lead to slightly different results.
- Future work:
    - Obtain positive result for every environment - is that even possible?
    - Find ways to speed up training or prevent local optima - adjust reward?

Annexes

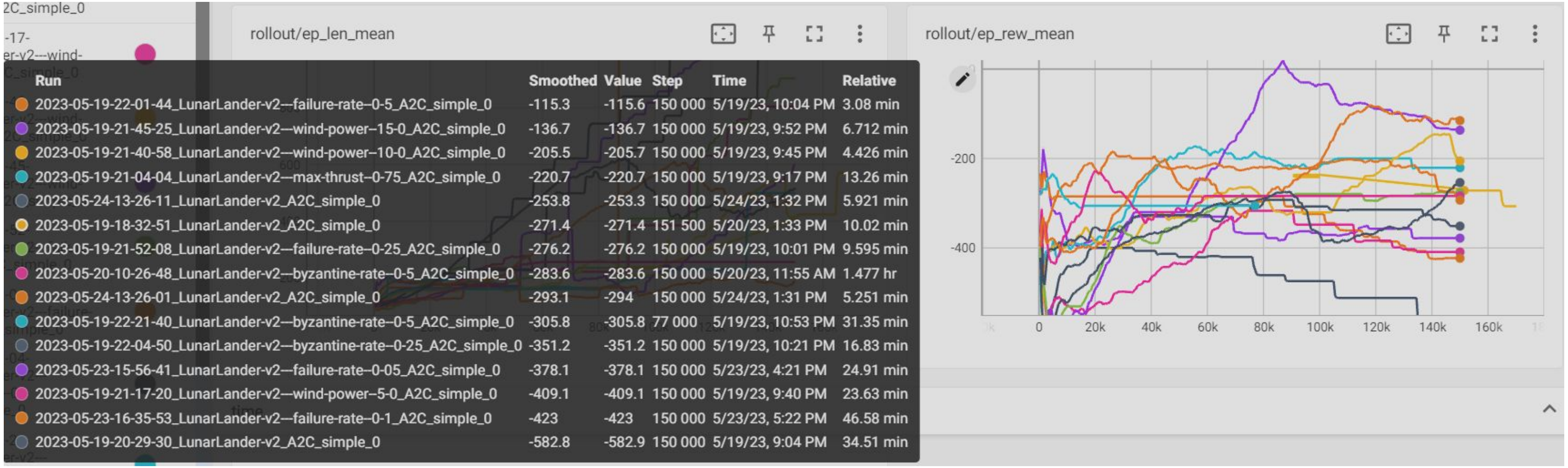
Note: The README.md in the code has instructions, other details on running, and the available features.

Other results

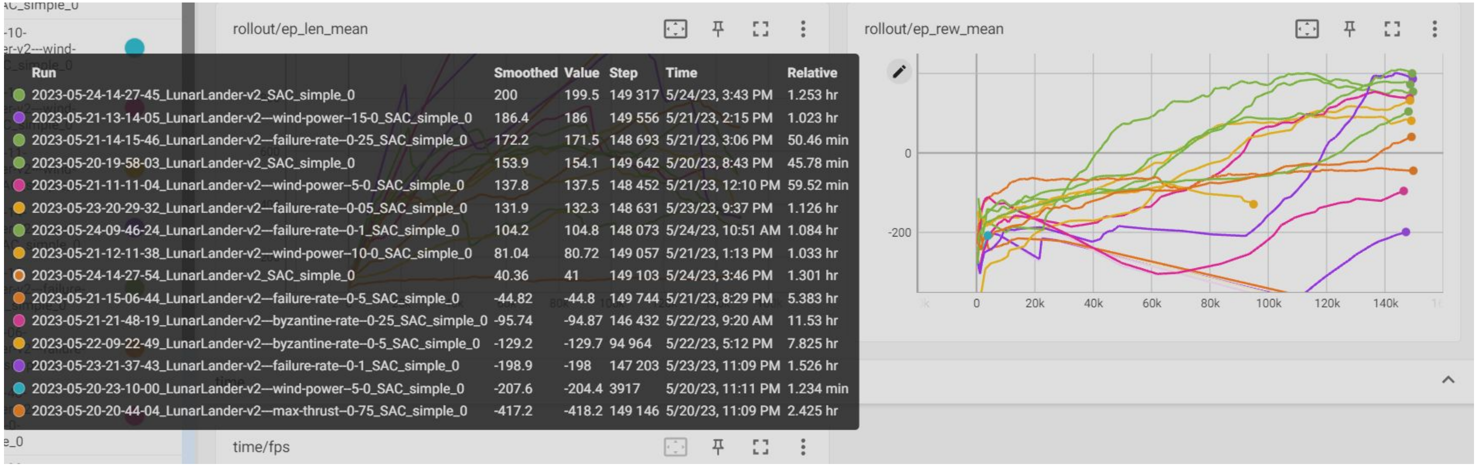
- Reward graphs per model



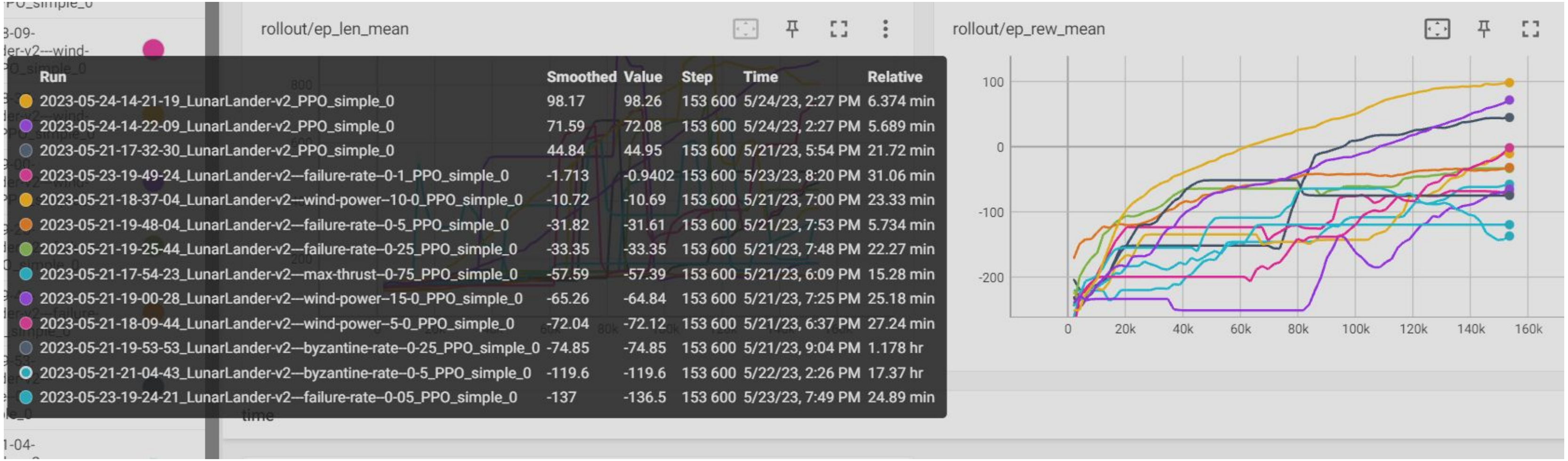
Run	Smoothed	Value	Step	Time	Relative
● 2023-05-24-13-31-22_LunarLander-v2_DDPG_simple_0	-21.41	-21.79	155 329	5/24/23, 2:21 PM	49.84 min
● 2023-05-24-13-32-10_LunarLander-v2_DDPG_simple_0	-88.72	-88.56	153 566	5/24/23, 2:21 PM	48.87 min
● 2023-05-20-17-30-23_LunarLander-v2_DDPG_simple_0	-141.7	-141.3	155 104	5/24/23, 10:32 AM	3.71 day



Run	Smoothed	Value	Step	Time	Relative
● 2023-05-19-22-01-44_LunarLander-v2---failure-rate--0-5_A2C_simple_0	-115.3	-115.6	150 000	5/19/23, 10:04 PM	3.08 min
● 2023-05-19-21-45-25_LunarLander-v2---wind-power--15-0_A2C_simple_0	-136.7	-136.7	150 000	5/19/23, 9:52 PM	6.712 min
● 2023-05-19-21-40-58_LunarLander-v2---wind-power--10-0_A2C_simple_0	-205.5	-205.7	150 000	5/19/23, 9:45 PM	4.426 min
● 2023-05-19-21-04-04_LunarLander-v2---max-thrust--0-75_A2C_simple_0	-220.7	-220.7	150 000	5/19/23, 9:17 PM	13.26 min
● 2023-05-24-13-26-11_LunarLander-v2_A2C_simple_0	-253.8	-253.3	150 000	5/24/23, 1:32 PM	5.921 min
● 2023-05-19-18-32-51_LunarLander-v2_A2C_simple_0	-271.4	-271.4	151 500	5/20/23, 1:33 PM	10.02 min
● 2023-05-19-21-52-08_LunarLander-v2---failure-rate--0-25_A2C_simple_0	-276.2	-276.2	150 000	5/19/23, 10:01 PM	9.595 min
● 2023-05-20-10-26-48_LunarLander-v2---byzantine-rate--0-5_A2C_simple_0	-283.6	-283.6	150 000	5/20/23, 11:55 AM	1.477 hr
● 2023-05-24-13-26-01_LunarLander-v2_A2C_simple_0	-293.1	-294	150 000	5/24/23, 1:31 PM	5.251 min
● 2023-05-19-22-21-40_LunarLander-v2---byzantine-rate--0-5_A2C_simple_0	-305.8	-305.8	77 000	5/19/23, 10:53 PM	31.35 min
● 2023-05-19-22-04-50_LunarLander-v2---byzantine-rate--0-25_A2C_simple_0	-351.2	-351.2	150 000	5/19/23, 10:21 PM	16.83 min
● 2023-05-23-15-56-41_LunarLander-v2---failure-rate--0-05_A2C_simple_0	-378.1	-378.1	150 000	5/23/23, 4:21 PM	24.91 min
● 2023-05-19-21-17-20_LunarLander-v2---wind-power--5-0_A2C_simple_0	-409.1	-409.1	150 000	5/19/23, 9:40 PM	23.63 min
● 2023-05-23-16-35-53_LunarLander-v2---failure-rate--0-1_A2C_simple_0	-423	-423	150 000	5/23/23, 5:22 PM	46.58 min
● 2023-05-19-20-29-30_LunarLander-v2_A2C_simple_0	-582.8	-582.9	150 000	5/19/23, 9:04 PM	34.51 min



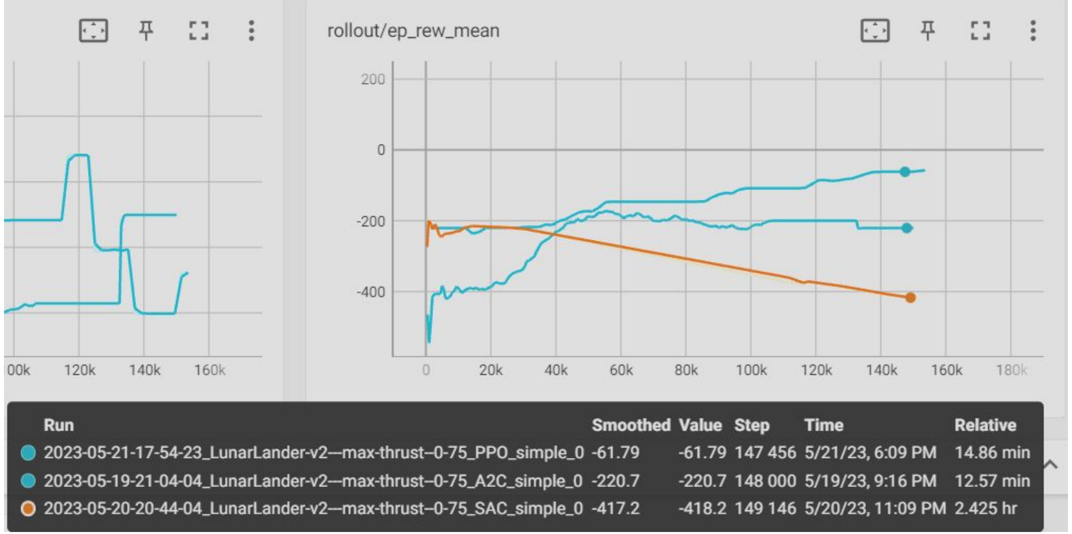
Run	Smoothed	Value	Step	Time	Relative
● 2023-05-24-14-27-45_LunarLander-v2_SAC_simple_0	200	199.5	149 317	5/24/23, 3:43 PM	1.253 hr
● 2023-05-21-13-14-05_LunarLander-v2—wind-power–15-0_SAC_simple_0	186.4	186	149 556	5/21/23, 2:15 PM	1.023 hr
● 2023-05-21-14-15-46_LunarLander-v2—failure-rate–0-25_SAC_simple_0	172.2	171.5	148 693	5/21/23, 3:06 PM	50.46 min
● 2023-05-20-19-58-03_LunarLander-v2_SAC_simple_0	153.9	154.1	149 642	5/20/23, 8:43 PM	45.78 min
● 2023-05-21-11-11-04_LunarLander-v2—wind-power–5-0_SAC_simple_0	137.8	137.5	148 452	5/21/23, 12:10 PM	59.52 min
● 2023-05-23-20-29-32_LunarLander-v2—failure-rate–0-05_SAC_simple_0	131.9	132.3	148 631	5/23/23, 9:37 PM	1.126 hr
● 2023-05-24-09-46-24_LunarLander-v2—failure-rate–0-1_SAC_simple_0	104.2	104.8	148 073	5/24/23, 10:51 AM	1.084 hr
● 2023-05-21-12-11-38_LunarLander-v2—wind-power–10-0_SAC_simple_0	81.04	80.72	149 057	5/21/23, 1:13 PM	1.033 hr
○ 2023-05-24-14-27-54_LunarLander-v2_SAC_simple_0	40.36	41	149 103	5/24/23, 3:46 PM	1.301 hr
● 2023-05-21-15-06-44_LunarLander-v2—failure-rate–0-5_SAC_simple_0	-44.82	-44.8	149 744	5/21/23, 8:29 PM	5.383 hr
● 2023-05-21-21-48-19_LunarLander-v2—byzantine-rate–0-25_SAC_simple_0	-95.74	-94.87	146 432	5/22/23, 9:20 AM	11.53 hr
● 2023-05-22-09-22-49_LunarLander-v2—byzantine-rate–0-5_SAC_simple_0	-129.2	-129.7	94 964	5/22/23, 5:12 PM	7.825 hr
● 2023-05-23-21-37-43_LunarLander-v2—failure-rate–0-1_SAC_simple_0	-198.9	-198	147 203	5/23/23, 11:09 PM	1.526 hr
● 2023-05-20-23-10-00_LunarLander-v2—wind-power–5-0_SAC_simple_0	-207.6	-204.4	3917	5/20/23, 11:11 PM	1.234 min
● 2023-05-20-20-44-04_LunarLander-v2—max-thrust–0-75_SAC_simple_0	-417.2	-418.2	149 146	5/20/23, 11:09 PM	2.425 hr



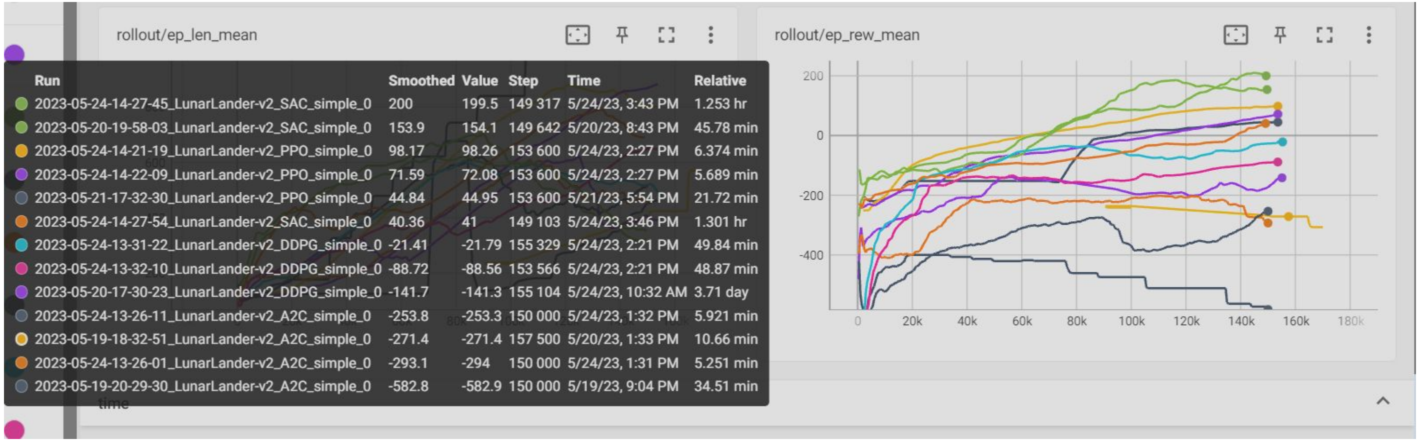
Run	Smoothed	Value	Step	Time	Relative
● 2023-05-24-14-21-19_LunarLander-v2_PPO_simple_0	98.17	98.26	153 600	5/24/23, 2:27 PM	6.374 min
● 2023-05-24-14-22-09_LunarLander-v2_PPO_simple_0	71.59	72.08	153 600	5/24/23, 2:27 PM	5.689 min
● 2023-05-21-17-32-30_LunarLander-v2_PPO_simple_0	44.84	44.95	153 600	5/21/23, 5:54 PM	21.72 min
● 2023-05-23-19-49-24_LunarLander-v2—failure-rate–0-1_PPO_simple_0	-1.713	-0.9402	153 600	5/23/23, 8:20 PM	31.06 min
● 2023-05-21-18-37-04_LunarLander-v2—wind-power–10-0_PPO_simple_0	-10.72	-10.69	153 600	5/21/23, 7:00 PM	23.33 min
● 2023-05-21-19-48-04_LunarLander-v2—failure-rate–0-5_PPO_simple_0	-31.82	-31.61	153 600	5/21/23, 7:53 PM	5.734 min
● 2023-05-21-19-25-44_LunarLander-v2—failure-rate–0-25_PPO_simple_0	-33.35	-33.35	153 600	5/21/23, 7:48 PM	22.27 min
● 2023-05-21-17-54-23_LunarLander-v2—max-thrust–0-75_PPO_simple_0	-57.59	-57.39	153 600	5/21/23, 6:09 PM	15.28 min
● 2023-05-21-19-00-28_LunarLander-v2—wind-power–15-0_PPO_simple_0	-65.26	-64.84	153 600	5/21/23, 7:25 PM	25.18 min
● 2023-05-21-18-09-44_LunarLander-v2—wind-power–5-0_PPO_simple_0	-72.04	-72.12	153 600	5/21/23, 6:37 PM	27.24 min
● 2023-05-21-19-53-53_LunarLander-v2—byzantine-rate–0-25_PPO_simple_0	-74.85	-74.85	153 600	5/21/23, 9:04 PM	1.178 hr
○ 2023-05-21-21-04-43_LunarLander-v2—byzantine-rate–0-5_PPO_simple_0	-119.6	-119.6	153 600	5/22/23, 2:26 PM	17.37 hr
● 2023-05-23-19-24-21_LunarLander-v2—failure-rate–0-05_PPO_simple_0	-137	-136.5	153 600	5/23/23, 7:49 PM	24.89 min

Other results

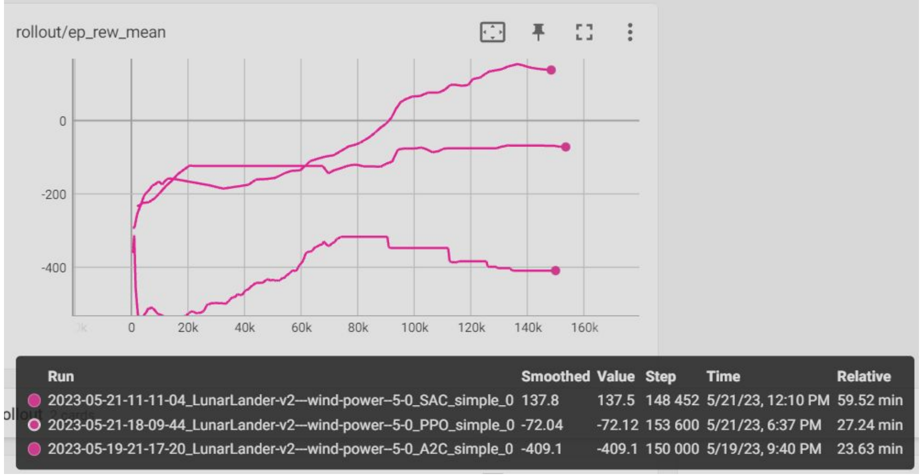
- Reward graphs per environment



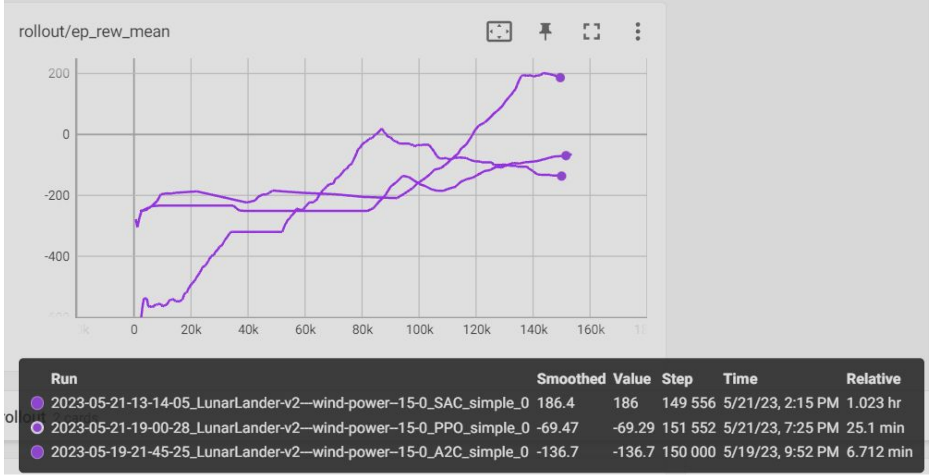
	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-21-17-54-23_LunarLander-v2—max-thrust–0-75_PPO_simple_0	-61.79	-61.79	147 456	5/21/23, 6:09 PM	14.86 min
●	2023-05-19-21-04-04_LunarLander-v2—max-thrust–0-75_A2C_simple_0	-220.7	-220.7	148 000	5/19/23, 9:16 PM	12.57 min
●	2023-05-20-20-44-04_LunarLander-v2—max-thrust–0-75_SAC_simple_0	-417.2	-418.2	149 146	5/20/23, 11:09 PM	2.425 hr



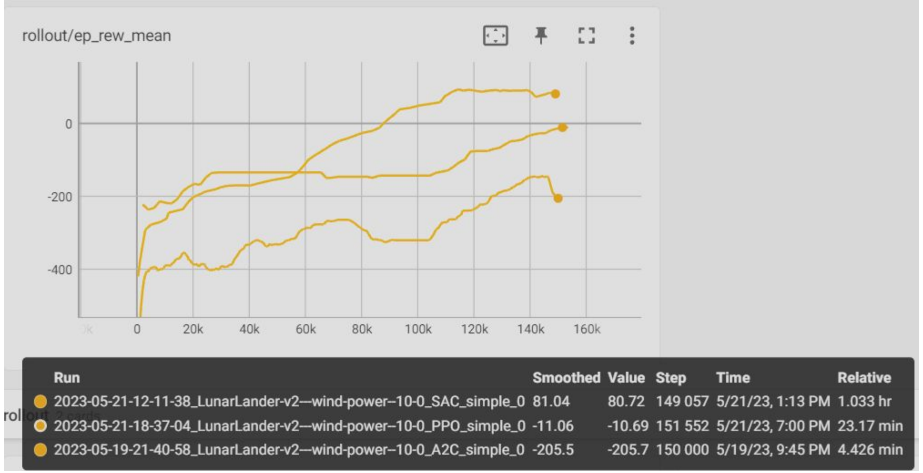
	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-24-14-27-45_LunarLander-v2_SAC_simple_0	200	199.5	149 317	5/24/23, 3:43 PM	1.253 hr
●	2023-05-20-19-58-03_LunarLander-v2_SAC_simple_0	153.9	154.1	149 642	5/20/23, 8:43 PM	45.78 min
●	2023-05-24-14-21-19_LunarLander-v2_PPO_simple_0	98.17	98.26	153 600	5/24/23, 2:27 PM	6.374 min
●	2023-05-24-14-22-09_LunarLander-v2_PPO_simple_0	71.59	72.08	153 600	5/24/23, 2:27 PM	5.689 min
●	2023-05-21-17-32-30_LunarLander-v2_PPO_simple_0	44.84	44.95	153 600	5/21/23, 5:54 PM	21.72 min
●	2023-05-24-14-27-54_LunarLander-v2_SAC_simple_0	40.36	41	149 103	5/24/23, 3:46 PM	1.301 hr
●	2023-05-24-13-31-22_LunarLander-v2_DDPG_simple_0	-21.41	-21.79	155 329	5/24/23, 2:21 PM	49.84 min
●	2023-05-24-13-32-10_LunarLander-v2_DDPG_simple_0	-88.72	-88.56	153 566	5/24/23, 2:21 PM	48.87 min
●	2023-05-20-17-30-23_LunarLander-v2_DDPG_simple_0	-141.7	-141.3	155 104	5/24/23, 10:32 AM	3.71 day
●	2023-05-24-13-26-11_LunarLander-v2_A2C_simple_0	-253.8	-253.3	150 000	5/24/23, 1:32 PM	5.921 min
●	2023-05-19-18-32-51_LunarLander-v2_A2C_simple_0	-271.4	-271.4	157 500	5/20/23, 1:33 PM	10.66 min
●	2023-05-24-13-26-01_LunarLander-v2_A2C_simple_0	-293.1	-294	150 000	5/24/23, 1:31 PM	5.251 min
●	2023-05-19-20-29-30_LunarLander-v2_A2C_simple_0	-582.8	-582.9	150 000	5/19/23, 9:04 PM	34.51 min



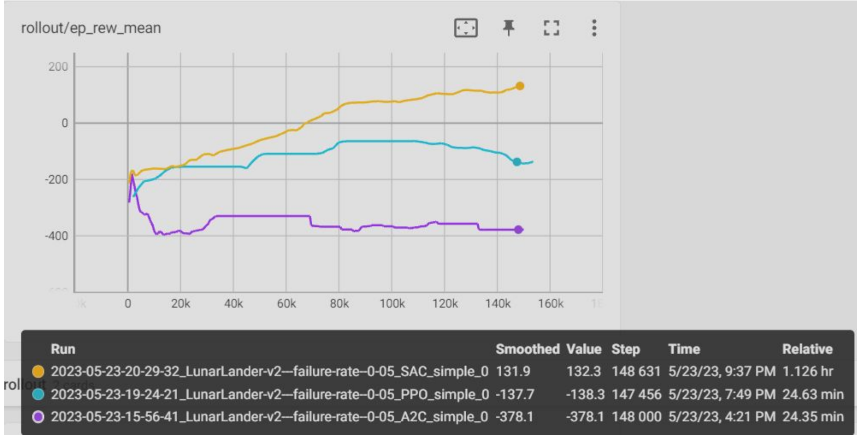
	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-21-11-11-04_LunarLander-v2—wind-power–5-0_SAC_simple_0	137.8	137.5	148 452	5/21/23, 12:10 PM	59.52 min
○	2023-05-21-18-09-44_LunarLander-v2—wind-power–5-0_PPO_simple_0	-72.04	-72.12	153 600	5/21/23, 6:37 PM	27.24 min
●	2023-05-19-21-17-20_LunarLander-v2—wind-power–5-0_A2C_simple_0	-409.1	-409.1	150 000	5/19/23, 9:40 PM	23.63 min



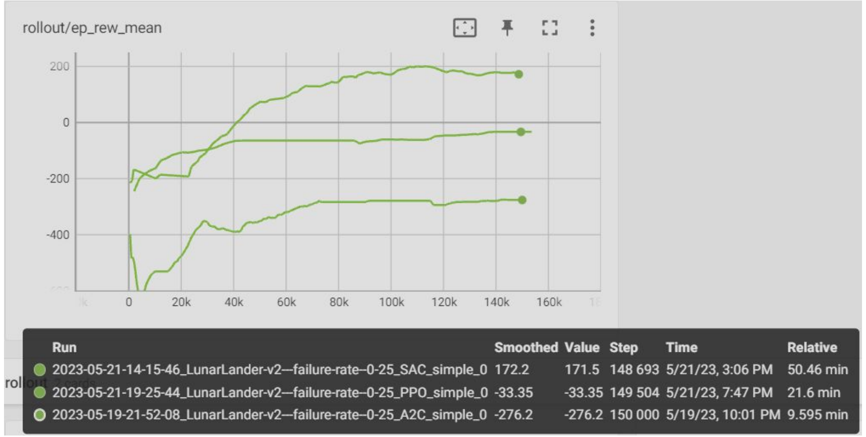
	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-21-13-14-05_LunarLander-v2—wind-power–15-0_SAC_simple_0	186.4	186	149 556	5/21/23, 2:15 PM	1.023 hr
○	2023-05-21-19-00-28_LunarLander-v2—wind-power–15-0_PPO_simple_0	-69.47	-69.29	151 552	5/21/23, 7:25 PM	25.1 min
●	2023-05-19-21-45-25_LunarLander-v2—wind-power–15-0_A2C_simple_0	-136.7	-136.7	150 000	5/19/23, 9:52 PM	6.712 min



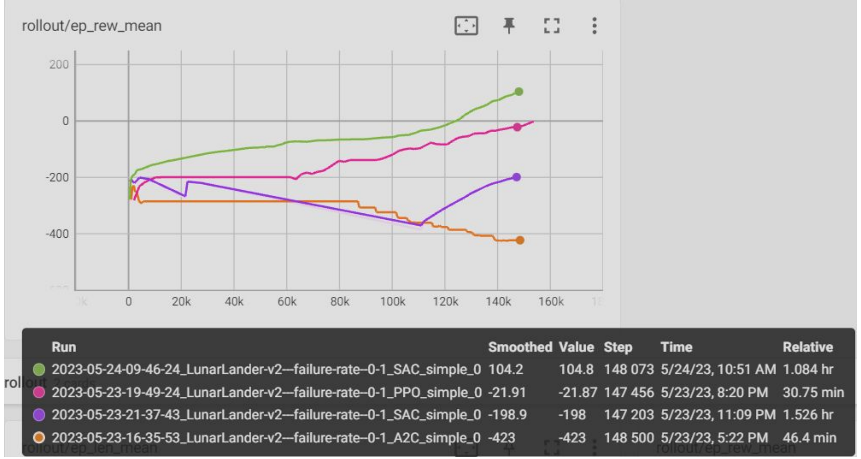
	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-21-12-11-38_LunarLander-v2—wind-power–10-0_SAC_simple_0	81.04	80.72	149 057	5/21/23, 1:13 PM	1.033 hr
○	2023-05-21-18-37-04_LunarLander-v2—wind-power–10-0_PPO_simple_0	-11.06	-10.69	151 552	5/21/23, 7:00 PM	23.17 min
●	2023-05-19-21-40-58_LunarLander-v2—wind-power–10-0_A2C_simple_0	-205.5	-205.7	150 000	5/19/23, 9:45 PM	4.426 min



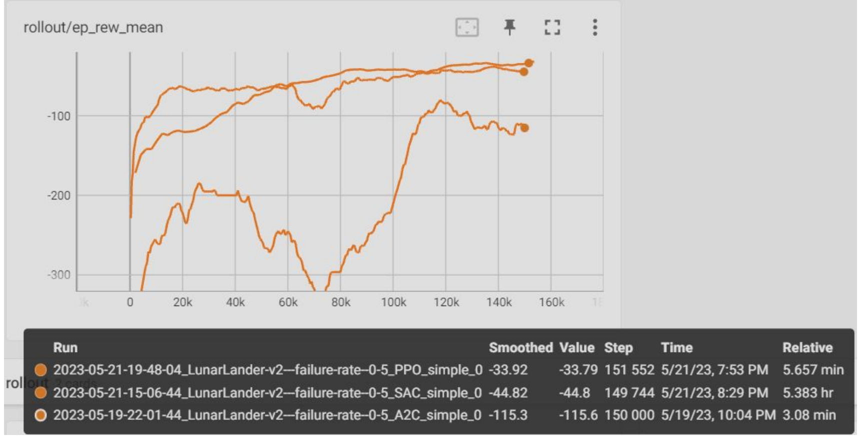
	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-23-20-29-32_LunarLander-v2—failure-rate–0-05_SAC_simple_0	131.9	132.3	148 631	5/23/23, 9:37 PM	1.126 hr
●	2023-05-23-19-24-21_LunarLander-v2—failure-rate–0-05_PPO_simple_0	-137.7	-138.3	147 456	5/23/23, 7:49 PM	24.63 min
●	2023-05-23-15-56-41_LunarLander-v2—failure-rate–0-05_A2C_simple_0	-378.1	-378.1	148 000	5/23/23, 4:21 PM	24.35 min



	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-21-14-15-46_LunarLander-v2—failure-rate–0-25_SAC_simple_0	172.2	171.5	148 693	5/21/23, 3:06 PM	50.46 min
●	2023-05-21-19-25-44_LunarLander-v2—failure-rate–0-25_PPO_simple_0	-33.35	-33.35	149 504	5/21/23, 7:47 PM	21.6 min
●	2023-05-19-21-52-08_LunarLander-v2—failure-rate–0-25_A2C_simple_0	-276.2	-276.2	150 000	5/19/23, 10:01 PM	9.595 min

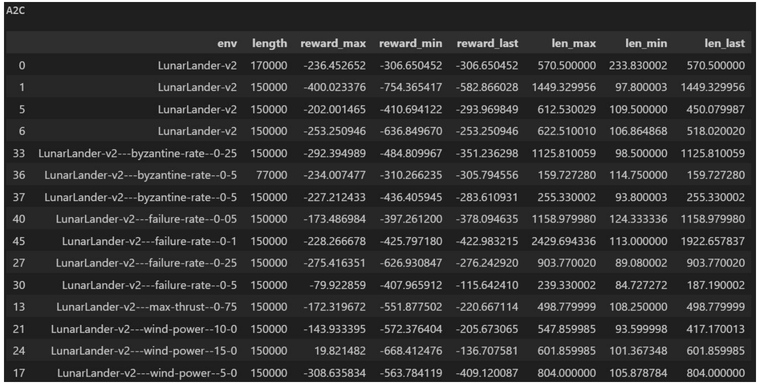


	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-24-09-46-24_LunarLander-v2—failure-rate–0-1_SAC_simple_0	104.2	104.8	148 073	5/24/23, 10:51 AM	1.084 hr
●	2023-05-23-19-49-24_LunarLander-v2—failure-rate–0-1_PPO_simple_0	-21.91	-21.87	147 456	5/23/23, 8:20 PM	30.75 min
●	2023-05-23-21-37-43_LunarLander-v2—failure-rate–0-1_SAC_simple_0	-198.9	-198	147 203	5/23/23, 11:09 PM	1.526 hr
●	2023-05-23-16-35-53_LunarLander-v2—failure-rate–0-1_A2C_simple_0	-423	-423	148 500	5/23/23, 5:22 PM	46.4 min



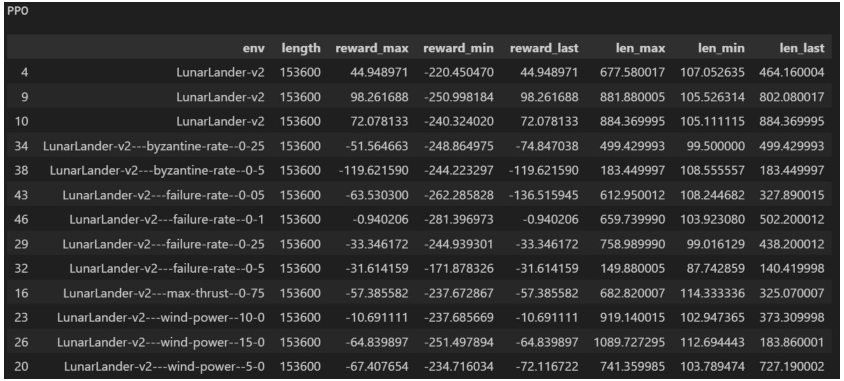
	Run	Smoothed	Value	Step	Time	Relative
●	2023-05-21-19-48-04_LunarLander-v2—failure-rate–0-5_PPO_simple_0	-33.92	-33.79	151 552	5/21/23, 7:53 PM	5.657 min
●	2023-05-21-15-06-44_LunarLander-v2—failure-rate–0-5_SAC_simple_0	-44.82	-44.8	149 744	5/21/23, 8:29 PM	5.383 hr
●	2023-05-19-22-01-44_LunarLander-v2—failure-rate–0-5_A2C_simple_0	-115.3	-115.6	150 000	5/19/23, 10:04 PM	3.08 min

Detailed data - By model



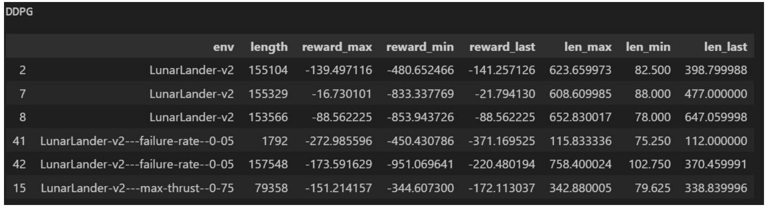
A2C

	env	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
0	LunarLander-v2	170000	-236.452652	-306.650452	-306.650452	570.500000	233.830002	570.500000
1	LunarLander-v2	150000	-400.023376	-754.365417	-582.866028	1449.329956	97.800003	1449.329956
5	LunarLander-v2	150000	-202.001465	-410.694122	-293.969849	612.530029	109.500000	450.079987
6	LunarLander-v2	150000	-253.250946	-636.849670	-253.250946	622.510010	106.864868	518.020020
33	LunarLander-v2---byzantine-rate--0-25	150000	-292.394989	-484.809967	-351.236298	1125.810059	98.500000	1125.810059
36	LunarLander-v2---byzantine-rate--0-5	77000	-234.007477	-310.266235	-305.794556	159.727280	114.750000	159.727280
37	LunarLander-v2---byzantine-rate--0-5	150000	-227.212433	-436.405945	-283.610931	255.330002	93.800003	255.330002
40	LunarLander-v2---failure-rate--0-05	150000	-173.486984	-397.261200	-378.094635	1158.979980	124.333336	1158.979980
45	LunarLander-v2---failure-rate--0-1	150000	-228.266678	-425.797180	-422.983215	2429.694336	113.000000	1922.657837
27	LunarLander-v2---failure-rate--0-25	150000	-275.416351	-626.930847	-276.242920	903.770020	89.080002	903.770020
30	LunarLander-v2---failure-rate--0-5	150000	-79.922859	-407.965912	-115.642410	239.330002	84.727272	187.190002
13	LunarLander-v2---max-thrust--0-75	150000	-172.319672	-551.877502	-220.667114	498.779999	108.250000	498.779999
21	LunarLander-v2---wind-power--10-0	150000	-143.933395	-572.376404	-205.673065	547.859985	93.599998	417.170013
24	LunarLander-v2---wind-power--15-0	150000	19.821482	-668.412476	-136.707581	601.859985	101.367348	601.859985
17	LunarLander-v2---wind-power--5-0	150000	-308.635834	-563.784119	-409.120087	804.000000	105.878784	804.000000



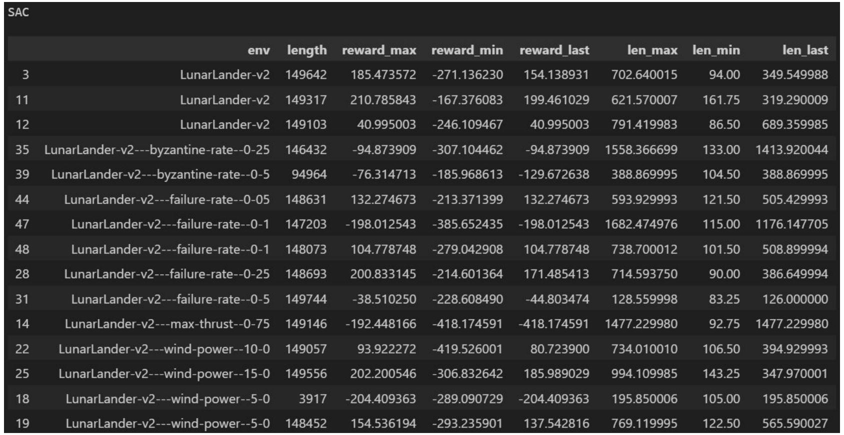
PPO

	env	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
4	LunarLander-v2	153600	44.948971	-220.450470	44.948971	677.580017	107.052635	464.160004
9	LunarLander-v2	153600	98.261688	-250.998184	98.261688	881.880005	105.526314	802.080017
10	LunarLander-v2	153600	72.078133	-240.324020	72.078133	884.369995	105.111115	884.369995
34	LunarLander-v2---byzantine-rate--0-25	153600	-51.564663	-248.864975	-74.847038	499.429993	99.500000	499.429993
38	LunarLander-v2---byzantine-rate--0-5	153600	-119.621590	-244.223297	-119.621590	183.449997	108.555557	183.449997
43	LunarLander-v2---failure-rate--0-05	153600	-63.530300	-262.285828	-136.515945	612.950012	108.244682	327.890015
46	LunarLander-v2---failure-rate--0-1	153600	-0.940206	-281.396973	-0.940206	659.739990	103.923080	502.200012
29	LunarLander-v2---failure-rate--0-25	153600	-33.346172	-244.939301	-33.346172	758.989990	99.016129	438.200012
32	LunarLander-v2---failure-rate--0-5	153600	-31.614159	-171.878326	-31.614159	149.880005	87.742859	140.419998
16	LunarLander-v2---max-thrust--0-75	153600	-57.385582	-237.672867	-57.385582	682.820007	114.333336	325.070007
23	LunarLander-v2---wind-power--10-0	153600	-10.691111	-237.685669	-10.691111	919.140015	102.947365	373.309998
26	LunarLander-v2---wind-power--15-0	153600	-64.839897	-251.497894	-64.839897	1089.727295	112.694443	183.860001
20	LunarLander-v2---wind-power--5-0	153600	-67.407654	-234.716034	-72.116722	741.359985	103.789474	727.190002



DDPG

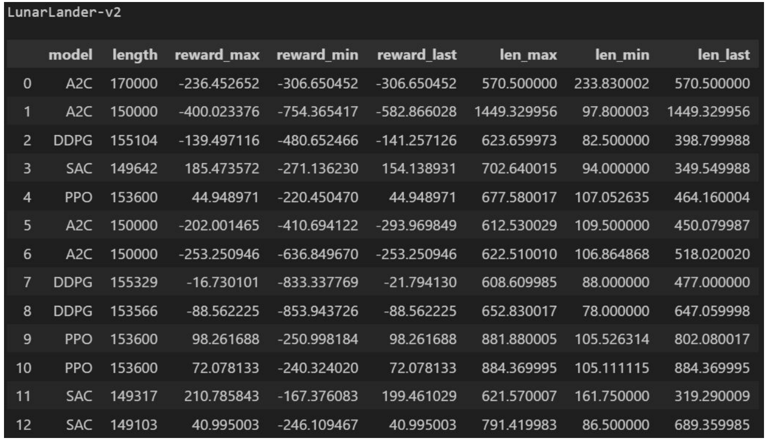
	env	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
2	LunarLander-v2	155104	-139.497116	-480.652466	-141.257126	623.659973	82.500	398.799988
7	LunarLander-v2	155329	-16.730101	-833.337769	-21.794130	608.609985	88.000	477.000000
8	LunarLander-v2	153566	-88.562225	-853.943726	-88.562225	652.830017	78.000	647.059998
41	LunarLander-v2---failure-rate--0-05	1792	-272.985596	-450.430786	-371.169525	115.833336	75.250	112.000000
42	LunarLander-v2---failure-rate--0-05	157548	-173.591629	-951.069641	-220.480194	758.400024	102.750	370.459991
15	LunarLander-v2---max-thrust--0-75	79358	-151.214157	-344.607300	-172.113037	342.880005	79.625	338.839996



SAC

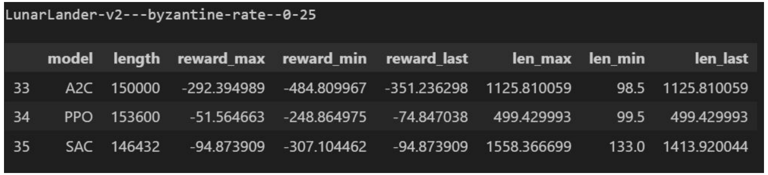
	env	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
3	LunarLander-v2	149642	185.473572	-271.136230	154.138931	702.640015	94.00	349.549988
11	LunarLander-v2	149317	210.785843	-167.376083	199.461029	621.570007	161.75	319.290009
12	LunarLander-v2	149103	40.995003	-246.109467	40.995003	791.419983	86.50	689.359985
35	LunarLander-v2---byzantine-rate--0-25	146432	-94.873909	-307.104462	-94.873909	1558.366699	133.00	1413.920044
39	LunarLander-v2---byzantine-rate--0-5	94964	-76.314713	-185.968613	-129.672638	388.869995	104.50	388.869995
44	LunarLander-v2---failure-rate--0-05	148631	132.274673	-213.371399	132.274673	593.929993	121.50	505.429993
47	LunarLander-v2---failure-rate--0-1	147203	-198.012543	-385.652435	-198.012543	1682.474976	115.00	1176.147705
48	LunarLander-v2---failure-rate--0-1	148073	104.778748	-279.042908	104.778748	738.700012	101.50	508.899994
28	LunarLander-v2---failure-rate--0-25	148693	200.833145	-214.601364	171.485413	714.593750	90.00	386.649994
31	LunarLander-v2---failure-rate--0-5	149744	-38.510250	-228.608490	-44.803474	128.559998	83.25	126.000000
14	LunarLander-v2---max-thrust--0-75	149146	-192.448166	-418.174591	-418.174591	1477.229980	92.75	1477.229980
22	LunarLander-v2---wind-power--10-0	149057	93.922272	-419.526001	80.723900	734.010010	106.50	394.929993
25	LunarLander-v2---wind-power--15-0	149556	202.200546	-306.832642	185.989029	994.109985	143.25	347.970001
18	LunarLander-v2---wind-power--5-0	3917	-204.409363	-289.090729	-204.409363	195.850006	105.00	195.850006
19	LunarLander-v2---wind-power--5-0	148452	154.536194	-293.235901	137.542816	769.119995	122.50	565.590027

Detailed data - By environment



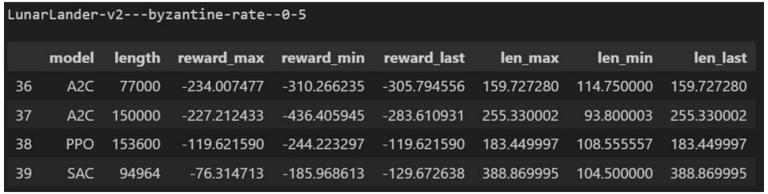
LunarLander-v2

	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
0	A2C	170000	-236.452652	-306.650452	-306.650452	570.500000	233.830002	570.500000
1	A2C	150000	-400.023376	-754.365417	-582.866028	1449.329956	97.800003	1449.329956
2	DDPG	155104	-139.497116	-480.652466	-141.257126	623.659973	82.500000	398.799988
3	SAC	149642	185.473572	-271.136230	154.138931	702.640015	94.000000	349.549988
4	PPO	153600	44.948971	-220.450470	44.948971	677.580017	107.052635	464.160004
5	A2C	150000	-202.001465	-410.694122	-293.969849	612.530029	109.500000	450.079987
6	A2C	150000	-253.250946	-636.849670	-253.250946	622.510010	106.864868	518.020020
7	DDPG	155329	-16.730101	-833.337769	-21.794130	608.609985	88.000000	477.000000
8	DDPG	153566	-88.562225	-853.943726	-88.562225	652.830017	78.000000	647.059998
9	PPO	153600	98.261688	-250.998184	98.261688	881.880005	105.526314	802.080017
10	PPO	153600	72.078133	-240.324020	72.078133	884.369995	105.111115	884.369995
11	SAC	149317	210.785843	-167.376083	199.461029	621.570007	161.750000	319.290009
12	SAC	149103	40.995003	-246.109467	40.995003	791.419983	86.500000	689.359985



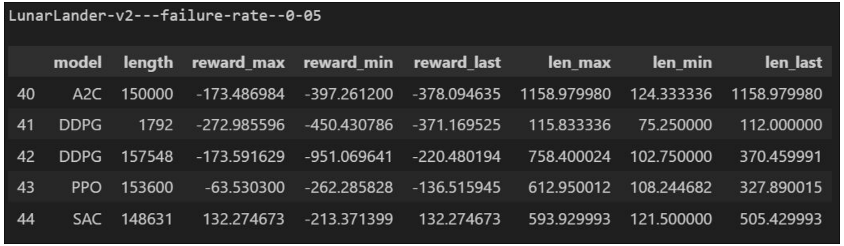
LunarLander-v2---byzantine-rate--0-25

	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
33	A2C	150000	-292.394989	-484.809967	-351.236298	1125.810059	98.5	1125.810059
34	PPO	153600	-51.564663	-248.864975	-74.847038	499.429993	99.5	499.429993
35	SAC	146432	-94.873909	-307.104462	-94.873909	1558.366699	133.0	1413.920044



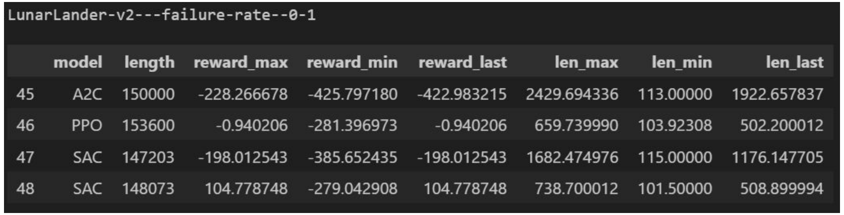
LunarLander-v2---byzantine-rate--0-5

	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
36	A2C	77000	-234.007477	-310.266235	-305.794556	159.727280	114.750000	159.727280
37	A2C	150000	-227.212433	-436.405945	-283.610931	255.330002	93.800003	255.330002
38	PPO	153600	-119.621590	-244.223297	-119.621590	183.449997	108.555557	183.449997
39	SAC	94964	-76.314713	-185.968613	-129.672638	388.869995	104.500000	388.869995



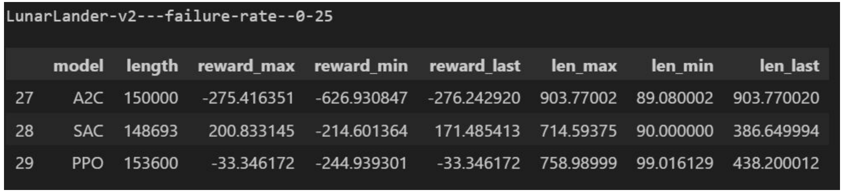
LunarLander-v2---failure-rate--0-05

	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
40	A2C	150000	-173.486984	-397.261200	-378.094635	1158.979980	124.333336	1158.979980
41	DDPG	1792	-272.985596	-450.430786	-371.169525	115.833336	75.250000	112.000000
42	DDPG	157548	-173.591629	-951.069641	-220.480194	758.400024	102.750000	370.459991
43	PPO	153600	-63.530300	-262.285828	-136.515945	612.950012	108.244682	327.890015
44	SAC	148631	132.274673	-213.371399	132.274673	593.929993	121.500000	505.429993



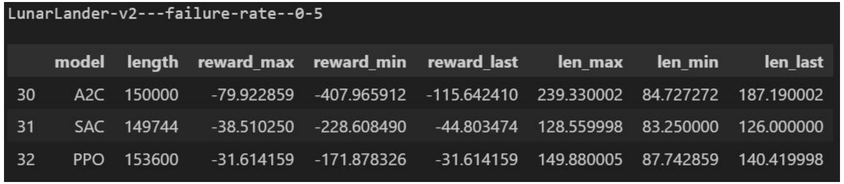
LunarLander-v2---failure-rate--0-1

	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
45	A2C	150000	-228.266678	-425.797180	-422.983215	2429.694336	113.00000	1922.657837
46	PPO	153600	-0.940206	-281.396973	-0.940206	659.739990	103.92308	502.200012
47	SAC	147203	-198.012543	-385.652435	-198.012543	1682.474976	115.00000	1176.147705
48	SAC	148073	104.778748	-279.042908	104.778748	738.700012	101.50000	508.899994



LunarLander-v2---failure-rate--0-25

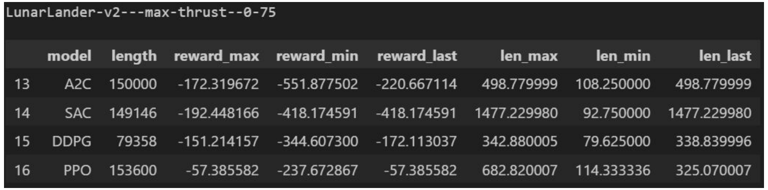
	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
27	A2C	150000	-275.416351	-626.930847	-276.242920	903.77002	89.080002	903.770020
28	SAC	148693	200.833145	-214.601364	171.485413	714.59375	90.000000	386.649994
29	PPO	153600	-33.346172	-244.939301	-33.346172	758.98999	99.016129	438.200012



LunarLander-v2---failure-rate--0-5

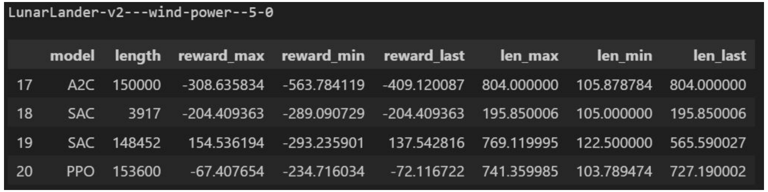
	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
30	A2C	150000	-79.922859	-407.965912	-115.642410	239.330002	84.727272	187.190002
31	SAC	149744	-38.510250	-228.608490	-44.803474	128.559998	83.250000	126.000000
32	PPO	153600	-31.614159	-171.878326	-31.614159	149.880005	87.742859	140.419998

Detailed data - By environment



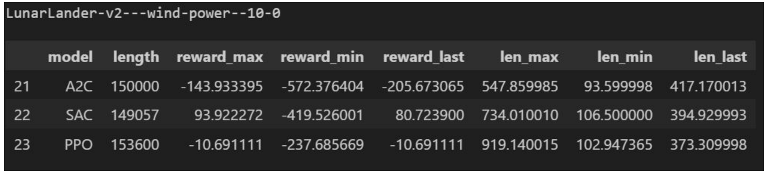
```
LunarLander-v2---max-thrust--0-75
```

	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
13	A2C	150000	-172.319672	-551.877502	-220.667114	498.779999	108.250000	498.779999
14	SAC	149146	-192.448166	-418.174591	-418.174591	1477.229980	92.750000	1477.229980
15	DDPG	79358	-151.214157	-344.607300	-172.113037	342.880005	79.625000	338.839996
16	PPO	153600	-57.385582	-237.672867	-57.385582	682.820007	114.333336	325.070007



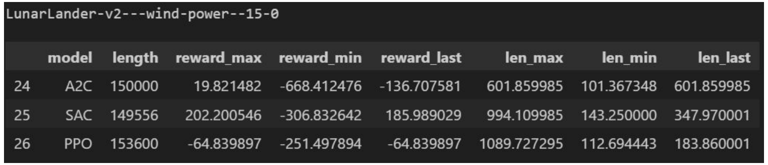
```
LunarLander-v2---wind-power--5-0
```

	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
17	A2C	150000	-308.635834	-563.784119	-409.120087	804.000000	105.878784	804.000000
18	SAC	3917	-204.409363	-289.090729	-204.409363	195.850006	105.000000	195.850006
19	SAC	148452	154.536194	-293.235901	137.542816	769.119995	122.500000	565.590027
20	PPO	153600	-67.407654	-234.716034	-72.116722	741.359985	103.789474	727.190002



```
LunarLander-v2---wind-power--10-0
```

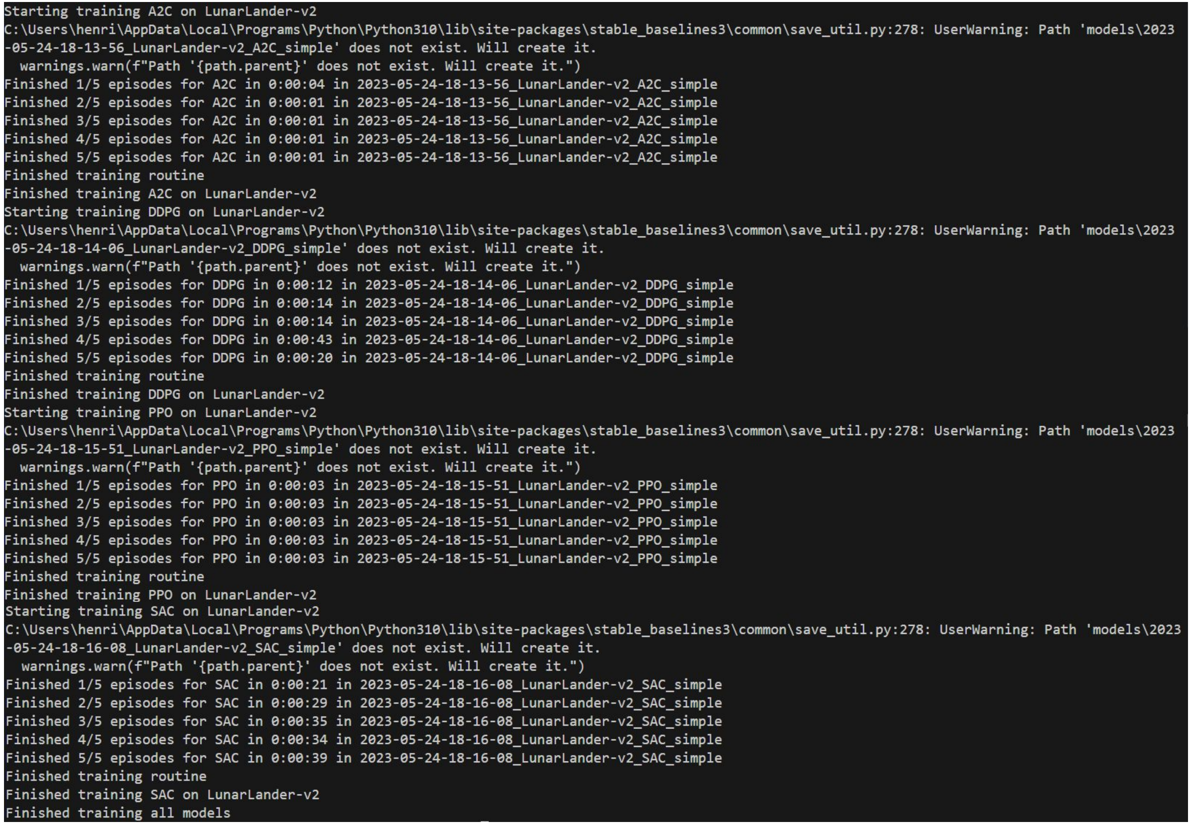
	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
21	A2C	150000	-143.933395	-572.376404	-205.673065	547.859985	93.599998	417.170013
22	SAC	149057	93.922272	-419.526001	80.723900	734.010010	106.500000	394.929993
23	PPO	153600	-10.691111	-237.685669	-10.691111	919.140015	102.947365	373.309998



```
LunarLander-v2---wind-power--15-0
```

	model	length	reward_max	reward_min	reward_last	len_max	len_min	len_last
24	A2C	150000	19.821482	-668.412476	-136.707581	601.859985	101.367348	601.859985
25	SAC	149556	202.200546	-306.832642	185.989029	994.109985	143.250000	347.970001
26	PPO	153600	-64.839897	-251.497894	-64.839897	1089.727295	112.694443	183.860001

Running all models



```
Starting training A2C on LunarLander-v2
C:\Users\henri\AppData\Local\Programs\Python\Python310\lib\site-packages\stable_baselines3\common\save_util.py:278: UserWarning: Path 'models\2023
-05-24-18-13-56_LunarLander-v2_A2C_simple' does not exist. Will create it.
  warnings.warn(f"Path '{path.parent}' does not exist. Will create it.")
Finished 1/5 episodes for A2C in 0:00:04 in 2023-05-24-18-13-56_LunarLander-v2_A2C_simple
Finished 2/5 episodes for A2C in 0:00:01 in 2023-05-24-18-13-56_LunarLander-v2_A2C_simple
Finished 3/5 episodes for A2C in 0:00:01 in 2023-05-24-18-13-56_LunarLander-v2_A2C_simple
Finished 4/5 episodes for A2C in 0:00:01 in 2023-05-24-18-13-56_LunarLander-v2_A2C_simple
Finished 5/5 episodes for A2C in 0:00:01 in 2023-05-24-18-13-56_LunarLander-v2_A2C_simple
Finished training routine
Finished training A2C on LunarLander-v2
Starting training DDPG on LunarLander-v2
C:\Users\henri\AppData\Local\Programs\Python\Python310\lib\site-packages\stable_baselines3\common\save_util.py:278: UserWarning: Path 'models\2023
-05-24-18-14-06_LunarLander-v2_DDPG_simple' does not exist. Will create it.
  warnings.warn(f"Path '{path.parent}' does not exist. Will create it.")
Finished 1/5 episodes for DDPG in 0:00:12 in 2023-05-24-18-14-06_LunarLander-v2_DDPG_simple
Finished 2/5 episodes for DDPG in 0:00:14 in 2023-05-24-18-14-06_LunarLander-v2_DDPG_simple
Finished 3/5 episodes for DDPG in 0:00:14 in 2023-05-24-18-14-06_LunarLander-v2_DDPG_simple
Finished 4/5 episodes for DDPG in 0:00:43 in 2023-05-24-18-14-06_LunarLander-v2_DDPG_simple
Finished 5/5 episodes for DDPG in 0:00:20 in 2023-05-24-18-14-06_LunarLander-v2_DDPG_simple
Finished training routine
Finished training DDPG on LunarLander-v2
Starting training PPO on LunarLander-v2
C:\Users\henri\AppData\Local\Programs\Python\Python310\lib\site-packages\stable_baselines3\common\save_util.py:278: UserWarning: Path 'models\2023
-05-24-18-15-51_LunarLander-v2_PPO_simple' does not exist. Will create it.
  warnings.warn(f"Path '{path.parent}' does not exist. Will create it.")
Finished 1/5 episodes for PPO in 0:00:03 in 2023-05-24-18-15-51_LunarLander-v2_PPO_simple
Finished 2/5 episodes for PPO in 0:00:03 in 2023-05-24-18-15-51_LunarLander-v2_PPO_simple
Finished 3/5 episodes for PPO in 0:00:03 in 2023-05-24-18-15-51_LunarLander-v2_PPO_simple
Finished 4/5 episodes for PPO in 0:00:03 in 2023-05-24-18-15-51_LunarLander-v2_PPO_simple
Finished 5/5 episodes for PPO in 0:00:03 in 2023-05-24-18-15-51_LunarLander-v2_PPO_simple
Finished training routine
Finished training PPO on LunarLander-v2
Starting training SAC on LunarLander-v2
C:\Users\henri\AppData\Local\Programs\Python\Python310\lib\site-packages\stable_baselines3\common\save_util.py:278: UserWarning: Path 'models\2023
-05-24-18-16-08_LunarLander-v2_SAC_simple' does not exist. Will create it.
  warnings.warn(f"Path '{path.parent}' does not exist. Will create it.")
Finished 1/5 episodes for SAC in 0:00:21 in 2023-05-24-18-16-08_LunarLander-v2_SAC_simple
Finished 2/5 episodes for SAC in 0:00:29 in 2023-05-24-18-16-08_LunarLander-v2_SAC_simple
Finished 3/5 episodes for SAC in 0:00:35 in 2023-05-24-18-16-08_LunarLander-v2_SAC_simple
Finished 4/5 episodes for SAC in 0:00:34 in 2023-05-24-18-16-08_LunarLander-v2_SAC_simple
Finished 5/5 episodes for SAC in 0:00:39 in 2023-05-24-18-16-08_LunarLander-v2_SAC_simple
Finished training routine
Finished training SAC on LunarLander-v2
Finished training all models
```