Lunar Lander Reinforcement Learning

Name: Soh Hong Yu

Admin Number: P2100775

Name: Samuel Tay Tze Ming

Admin Number: P2107404

Class: DAAA/FT/2B/01

Module Code: ST1504 Deep Learning

References (In Harvard format):

1. OpenAI (2022) Gymnasium documentation, Lunar Lander - Gymnasium Documentation.
   Available at: https://gymnasium.farama.org/environments/box2d/lunar_lander/ (Accessed: February 5, 2023).
2. van Hasselt, H., Guez, A. and Silver, D. (2015) Deep reinforcement learning with double Q-learning, arXiv.org.
   Available at: https://arxiv.org/abs/1509.06461 (Accessed: February 5, 2023).
3. Huang, S. et al. (2022) A2c is a special case of PPO, arXiv.org.
   Available at: https://arxiv.org/abs/2205.09123 (Accessed: February 5, 2023).
4. Schulman, J. et al. (2017) Proximal policy optimization algorithms, arXiv.org.
   Available at: https://arxiv.org/abs/1707.06347 (Accessed: February 5, 2023).
5. Adam, P. and Mark, T. (2021) Reinforcement learning (DQN) tutorial¶, Reinforcement Learning (DQN) Tutorial - PyTorch Tutorials 1.13.1+cu117 documentation.
   Available at: https://pytorch.org/tutorials/intermediate/reinforcement_q_learning.html (Accessed: February 5, 2023).
6. Tilbe, A. (2022) Actor-critic algorithm, Simplified: Essential for Finance and financial engineering, Medium. Level Up Coding.
   Available at: https://levelup.gitconnected.com/actor-critic-algorithm-simplified-essential-for-finance-and-financial-engineering-3ebc9ec78467 (Accessed: February 5, 2023).
7. Wang, M. (2021) Advantage actor critic tutorial: MINA2C, Medium. Towards Data Science.
   Available at: https://towardsdatascience.com/advantage-actor-critic-tutorial-mina2c-7a3249962fc8 (Accessed: February 5, 2023).
8. Hui, J. (2018) RL - trust region policy optimization (TRPO) explained, Medium. Medium.
   Available at: https://medium.com/@jonathan_hui/rl-trust-region-policy-optimization-trpo-explained-a6ee04eeeee9 (Accessed: February 5, 2023).
9. freeCodeCamp.org (2018) An intro to advantage actor critic methods: Let's play sonic the hedgehog!, freeCodeCamp.org. freeCodeCamp.org.
   Available at: https://www.freecodecamp.org/news/an-intro-to-advantage-actor-critic-methods-lets-play-sonic-the-hedgehog-86d6240171d/ (Accessed: February 5, 2023).
10. Yoon, C. (2019) Understanding actor critic methods, Medium. Towards Data Science.
    Available at: https://towardsdatascience.com/understanding-actor-critic-methods-931b97b6df3f (Accessed: February 5, 2023).

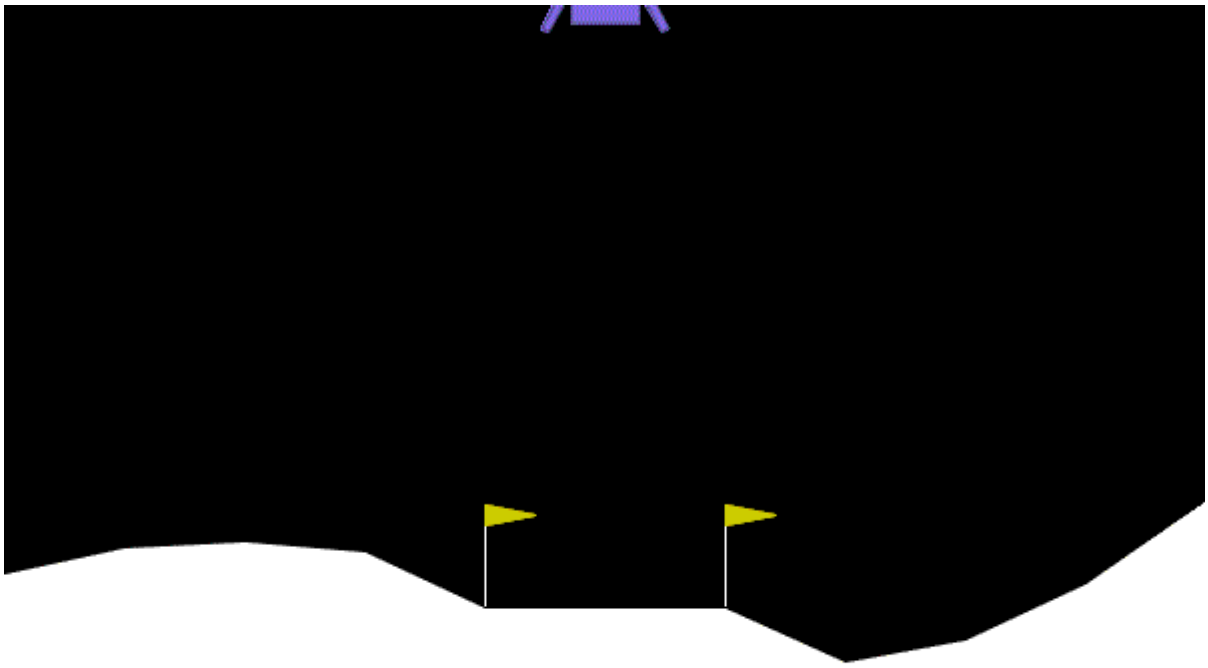
11. Levine, S. (2021) Actor-critic Algorithms - University of California, Berkeley, Actor-Critic Algorithms. Available at: http://rail.eecs.berkeley.edu/deeprlcourse-fa17/f17docs/lecture_5_actor_critic_pdf.pdf?source=post_page---------- (Accessed: February 5, 2023).
12. Lisi, A. (2021) Beating pong using reinforcement learning - part 2 A2c and PPO, Medium. Analytics Vidhya. Available at: https://medium.com/analytics-vidhya/beating-pong-using-reinforcement-learning-part-2-a2c-and-ppo-b83391dd3657 (Accessed: February 5, 2023).

Project Objective

Implement a suitable RL architecture to the problem. Land the LunarLander successfully on the landing pad.

Background Information

Lunar Lander is an environment is part of the Box2D environments.



This environment is a classic rocket trajectory optimization problem. According to Pontryagin's maximum principle, it is optimal to fire the engine at full throttle or turn it off. This is the reason why this environment has discrete actions: engine on or off.

There are two environment versions: discrete or continuous. The landing pad is always at coordinates (0,0). The coordinates are the first two numbers in the state vector. Landing outside of the landing pad is possible. Fuel is infinite, so an agent can learn to fly and then land on its first attempt.

There are a total of 4 discrete actions that the lander can do:

1. Do nothing
2. Fire left orientation engine
3. Fire main engine
4. Fire right orientation engine

There are a total of 8 observation space:

1. The coordinates of the lander in x
2. The coordinates of the lander in y
3. Its linear velocities in x
4. Its linear velocities in y
5. Its angle
6. Its angular velocity
7. If left leg is in contact with the ground
8. If right leg is in contact with the ground

After every step a reward is granted. The total reward of an episode is the sum of the rewards for all the steps within that episode.

For each step, the reward:

- is increased/decreased the closer/further the lander is to the landing pad.
- is increased/decreased the slower/faster the lander is moving.
- is decreased the more the lander is tilted (angle not horizontal).
- is increased by 10 points for each leg that is in contact with the ground.
- is decreased by 0.03 points each frame a side engine is firing.
- is decreased by 0.3 points each frame the main engine is firing.

The episode receive an additional reward of -100 or +100 points for crashing or landing safely respectively. An episode is considered a solution if it scores at least 200 points.

The lander starts at the top center of the viewport with a random initial force applied to its center of mass. The episode finishes if:

1. the lander crashes (the lander body gets in contact with the moon);
2. the lander gets outside of the viewport (x coordinate is greater than 1);
3. the lander is not awake. From the Box2D docs, a body which is not awake is a body which doesn't move and doesn't collide with any other body:

Initialising Setup

In [ ]:
```
# !apt-get install x11-utils > /dev/null 2>&1
# !pip install pyglet > /dev/null 2>&1
# !apt-get install -y xvfb python-opengl > /dev/null 2>&1
```

In [ ]:
```
# !pip install gym==0.25.2
# !pip install pyvirtualdisplay > /dev/null 2>&1
# !pip install gymnasium[box2d]
# !pip install torch==1.13.1
```

Initialising Libraries and Variables

In [ ]:
```
import gym
import numpy as np
import matplotlib.pyplot as plt
from IPython import display as ipythondisplay
from pyvirtualdisplay.display import Display
from IPython.display import clear_output, display
```

```
import torch
import torch.nn as nn
import torch.optim as optim
import torch.nn.functional as F
import torch.distributions as distributions
import random
import copy
import time
from matplotlib import animation, rc
from collections import deque, namedtuple
```

Checking GPU

In [ ]:
```
torch.cuda.is_available()
```

Out[ ]:
```
True
```

In [ ]:
```
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
print(device)
```

```
cuda:0
```

Loading Environment

Using the gym environment from gymnasium, we will make the LunarLander-v2 environment.

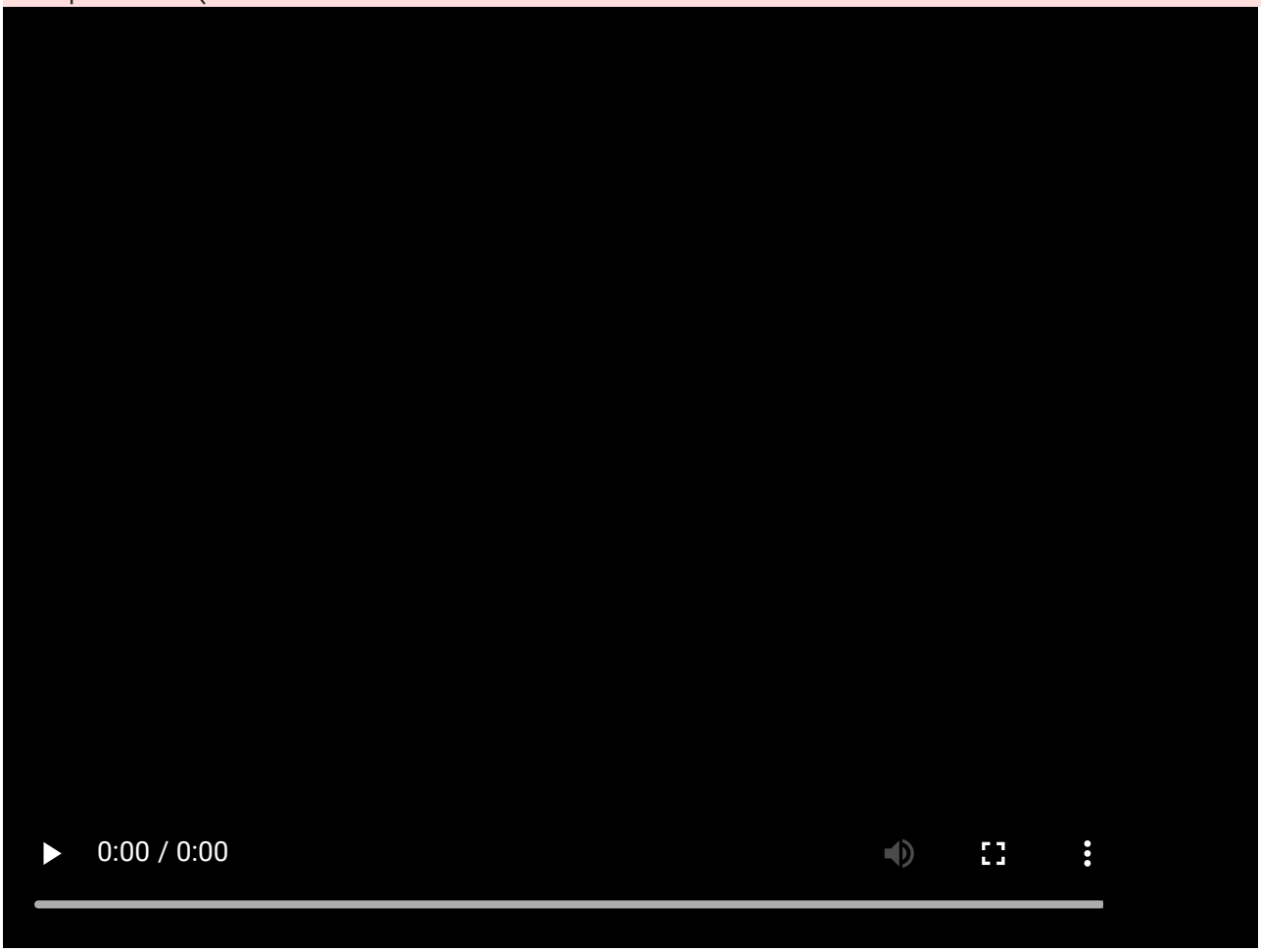
In [ ]:
```
def create_animation(frames, filename=None):
    rc("animation", html="jshtml")
    fig = plt.figure()
    plt.axis("off")
    im = plt.imshow(frames[0], animated=True)

    def updatefig(i):
        im.set_array(frames[i])
        return im,

    animationFig = animation.FuncAnimation(fig, updatefig, frames=len(frames), interval=len(f
    ipythondisplay.display(ipythondisplay.HTML(animationFig.to_html5_video()))
    if filename != None:
        animationFig.save(filename)
    return animationFig
```

Display environment by running 50 steps.

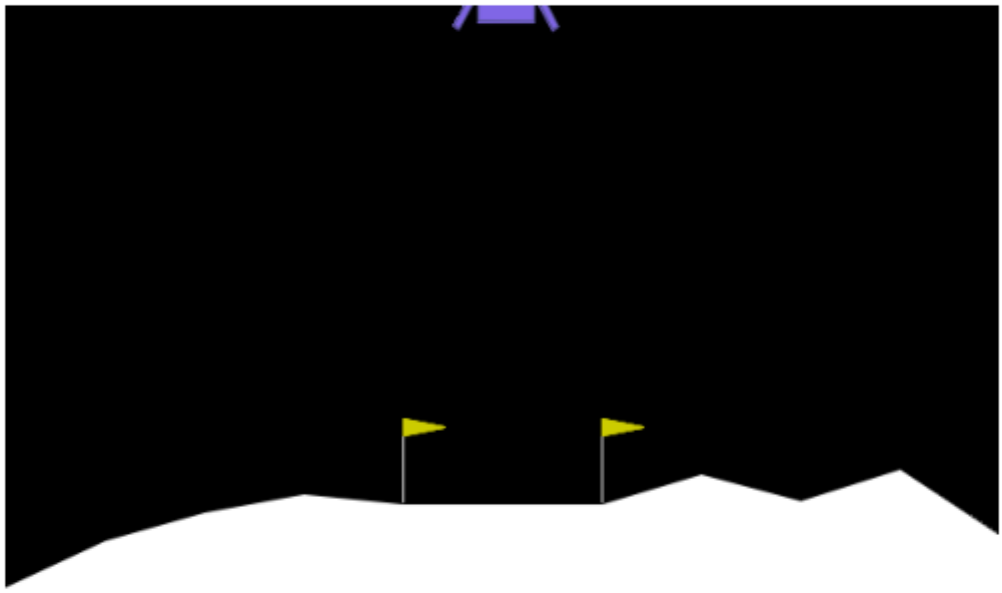
In [ ]:
```
env = gym.make(
    "LunarLander-v2",
    continuous = False,
    gravity = -10.0,
    enable_wind = False,
    wind_power = 15.0,
    turbulence_power = 1.5
)
env.action_space.seed(42)
env.reset()

frames = []

for i in range(50):
    action = env.action_space.sample()
    obs, reward, done, info = env.step(action)
    screen = env.render(mode='rgb_array')
    frames.append(screen)
```

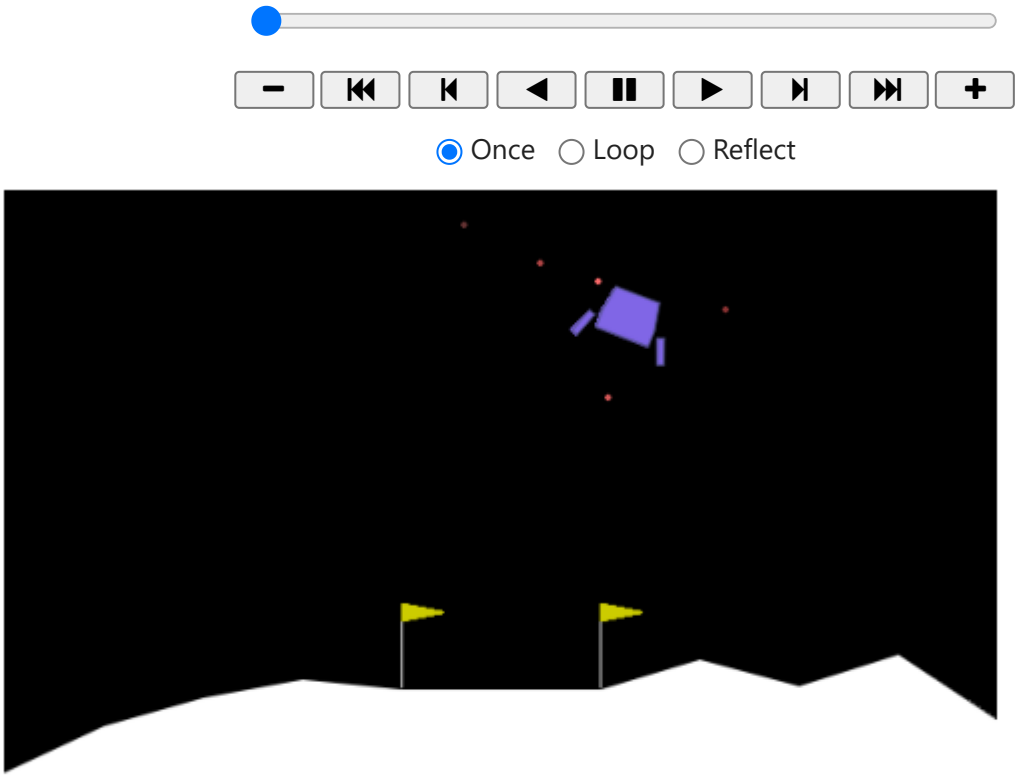
```
    if done:
        break

env.close()

create_animation(frames)
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:317: DeprecationWarning: WARN: Initializing wrapper in old step API which returns one bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\wrappers\step_api_compatibility.py:39: DeprecationWarning: WARN: Initializing environment in old step API which returns one bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\utils\passive_env_checker.py:241: DeprecationWarning: `np.bool8` is a deprecated alias for `np.bool_`.  (Deprecated NumPy 1.24)
  if not isinstance(terminated, (bool, np.bool8)):
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```



Out[ ]:





Exploratory Data Analysis

We will begin by conducting an exploratory data analysis of the environment to help us better understand the different actions and what they do to affect the lunar lander.

In [ ]:
```
print('Number of Observation Space: ', env.observation_space.shape)
print('Number of Actions: ', env.action_space)
```

We can see that as stated in the background information, there are 8 observation spaces and 4 possible actions.

Testing all actions and how it affects the lunar lander

First we want to understand what each action in the lunar lander do and its effects on the lunar lander.

Action 1 [Do Nothing]

In [ ]:
```
env = gym.make(
    "LunarLander-v2",
    continuous = False,
    gravity = -10.0,
    enable_wind = False,
    wind_power = 15.0,
    turbulence_power = 1.5
)
env.action_space.seed(42)
env.reset()

done = False

frames = []

while not done:
    obs, reward, done, info = env.step(0)
    screen = env.render(mode='rgb_array')
    frames.append(screen)
    if done:
        break

env.close()

create_animation(frames)
```

Observation

We note that by doing nothing, the lunar lander move based on the wind direction and the force exerted by the space itself. We note that the by doing nothing, there is no penalty applied which will affect the score which can help reduce the score to allow the lunar lander to get a higher score.

Action 2 [Left Engine Fire]

In [ ]:
```
env = gym.make(
    "LunarLander-v2",
    continuous = False,
    gravity = -10.0,
    enable_wind = False,
    wind_power = 15.0,
    turbulence_power = 1.5
)
env.action_space.seed(42)
env.reset()

done = False

frames = []

while not done:
    obs, reward, done, info = env.step(1)
    screen = env.render(mode='rgb_array')
    frames.append(screen)
    if done:
```

```
        break

env.close()

create_animation(frames)
```

Observation

We note that by activating left engine, the lunar lander will spin anti-clockwise. We need to take note that the by activating the side engine, there is a penalty of 0.03 applied to the overall score. We also note that this could be use to help stablise the lander if the wind is too strong etc.

Action 3 [Main Engine Fire]

In [ ]:
```
env = gym.make(
    "LunarLander-v2",
    continuous = False,
    gravity = -10.0,
    enable_wind = False,
    wind_power = 15.0,
    turbulence_power = 1.5
)
env.action_space.seed(42)
env.reset()

done = False

frames = []

while not done:
    obs, reward, done, info = env.step(2)
    screen = env.render(mode='rgb_array')
    frames.append(screen)
    if done:
        break
env.close()

create_animation(frames)
```

Observation

We note that by activating main engine, the lunar lander will move up. We need to take note that the by activating the main engine, there is a penalty of 0.3 applied to the overall score. We also note that this could be use to help gain height for the lander so that there is space for corrections to take place.

Action 4 [Right Engine Fire]

In [ ]:
```
env = gym.make(
    "LunarLander-v2",
    continuous = False,
    gravity = -10.0,
    enable_wind = False,
    wind_power = 15.0,
    turbulence_power = 1.5
)
env.action_space.seed(42)
env.reset()

done = False

frames = []

while not done:
    obs, reward, done, info = env.step(3)
```

```
    screen = env.render(mode='rgb_array')
    frames.append(screen)
    if done:
        break
env.close()

create_animation(frames)
```

Observation

We note that by activating right engine, the lunar lander will spin clockwise. We need to take note that the by activating the side engine, there is a penalty of 0.03 applied to the overall score. We also note that this could be use to help stablise the lander if the wind is too strong etc.

Summary

As main engine will allow the lunar lander to move upwards, this will allow the lunar lander to recorrect the position to allow it to land safely. By using the left and right engine, it allows the lunar lander to turn and spin to correct itself. As there is penalties for firing the engine [Simulate limited fuel], doing nothing allows the lander to go with the flow of the wind and reducing the penalties.

Building Models

We will be building a few reinforcement learning models to help land the lunar lander safely.

Model List:

1. Random Action Model (Baseline)
2. Deep Q Network (DQN)
3. Advantage Actor Critic (A2C)
4. Proximal Policy Optimization (PPO)

Random Action Model (Baseline)

Creating an agent that takes random actions

In [ ]:
```
max_episodes = 100
```

Setup Training

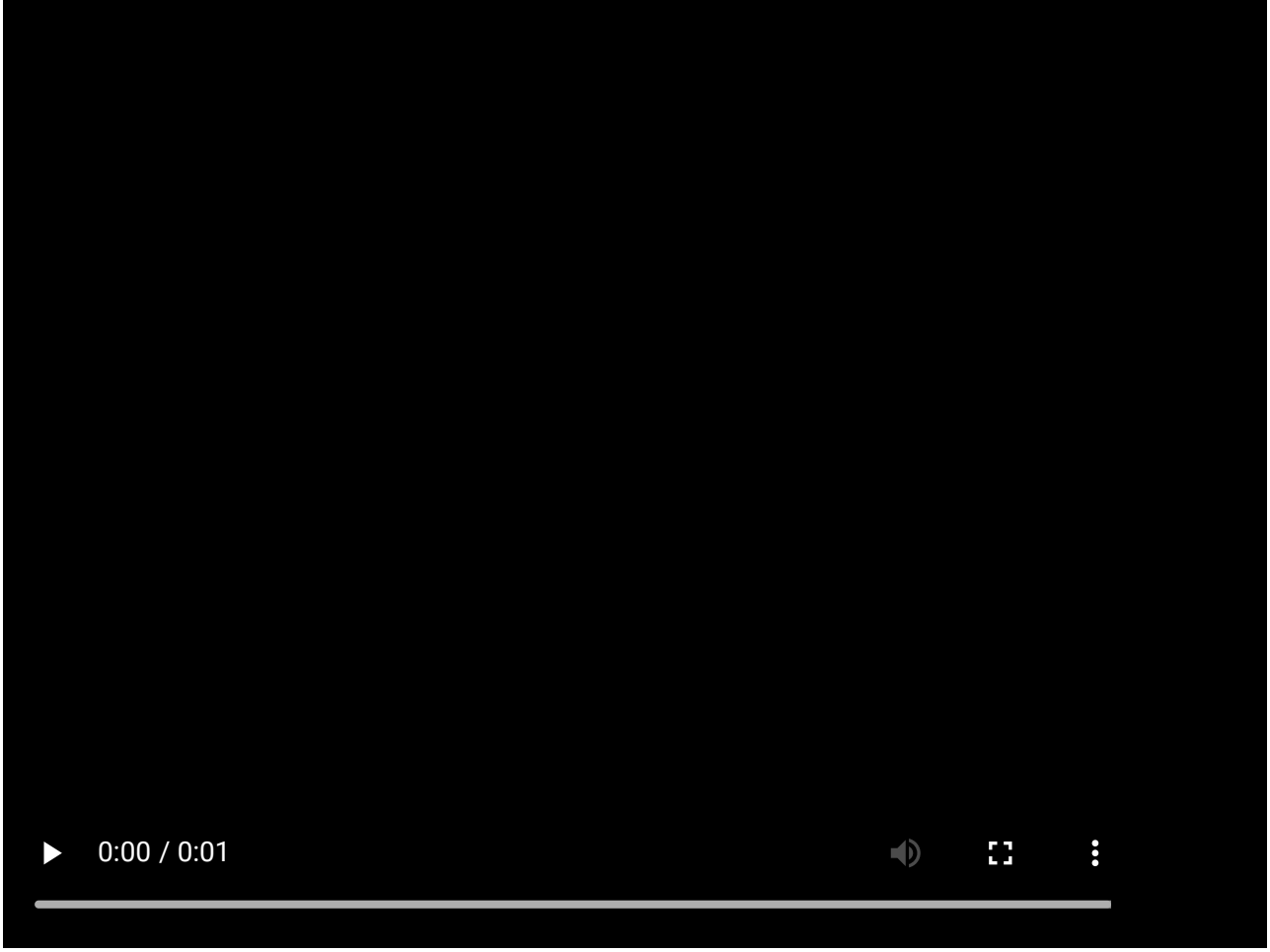
In [ ]:
```
env = gym.make(
    "LunarLander-v2",
    continuous = False,
    gravity = -10.0,
    enable_wind = False,
    wind_power = 15.0,
    turbulence_power = 1.5
)
env.action_space.seed(42)
env.reset()

scores = []
master_frames = []
actions = range(env.action_space.n)
for i in range(1, max_episodes+1):
    state = env.reset()
    frames = []
    score = 0
    while True:
```

```
        action = np.random.choice(actions)
        obs, reward, terminated, info = env.step(action)
        score += reward
        screen = env.render(mode='rgb_array')
        frames.append(screen)
        if terminated:
            if i % 20 == 0:
                create_animation(frames)
                print('Episode {},  score: {}'.format(i, score))
            break
    master_frames.append(frames)
    scores.append(score)
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:317: DeprecationWarning: WARN: Initializing wrapper in old step API which returns one bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\wrappers\step_api_compatibility.py:39: DeprecationWarning: WARN: Initializing environment in old step API which returns one bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\utils\passive_env_checker.py:241: DeprecationWarning: `np.bool8` is a deprecated alias for `np.bool_`.  (Deprecated NumPy 1.24)
  if not isinstance(terminated, (bool, np.bool8)):
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

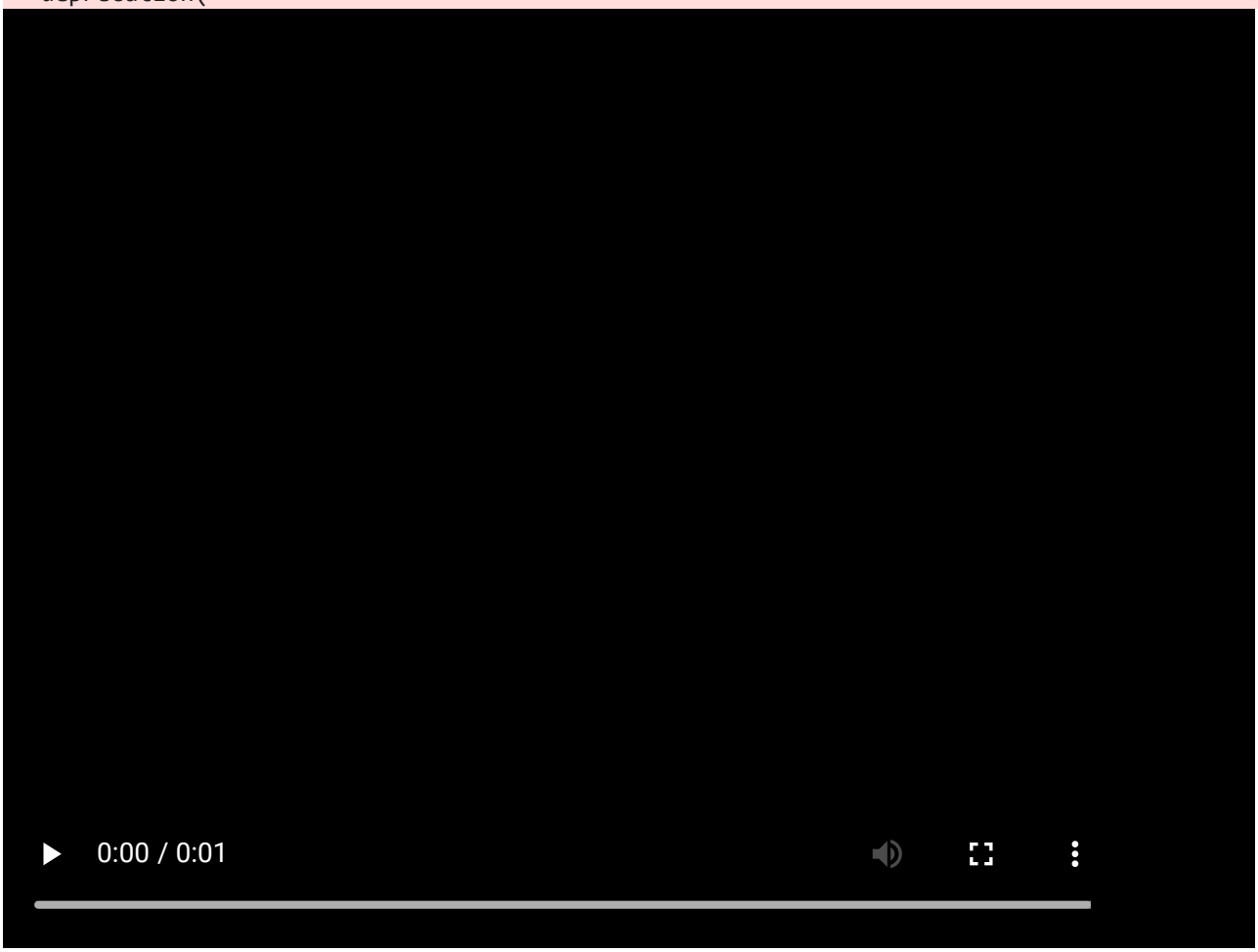

```
Episode 20,  score: -109.17129768506236
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```



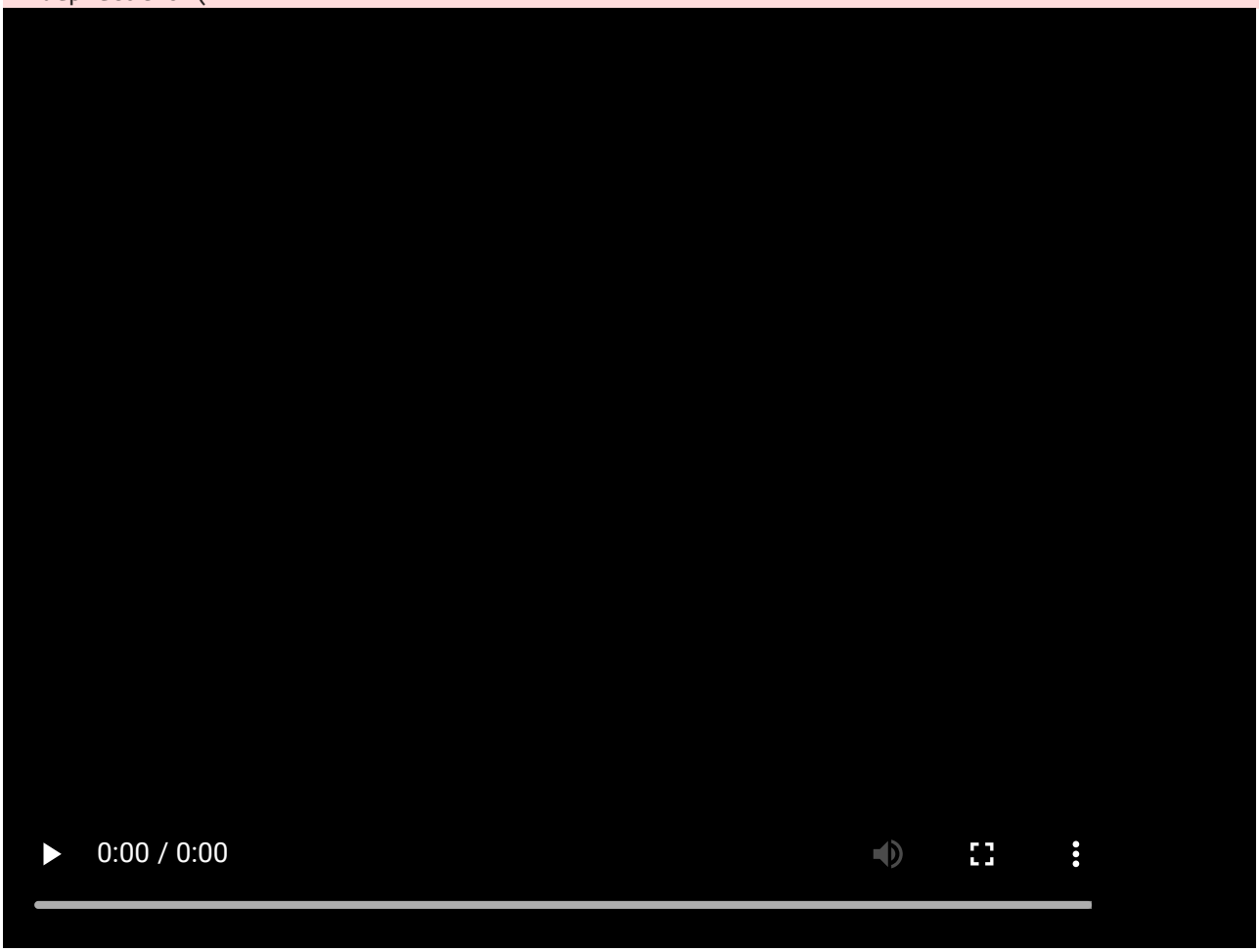

```
Episode 40,  score: -146.02518611543834
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```



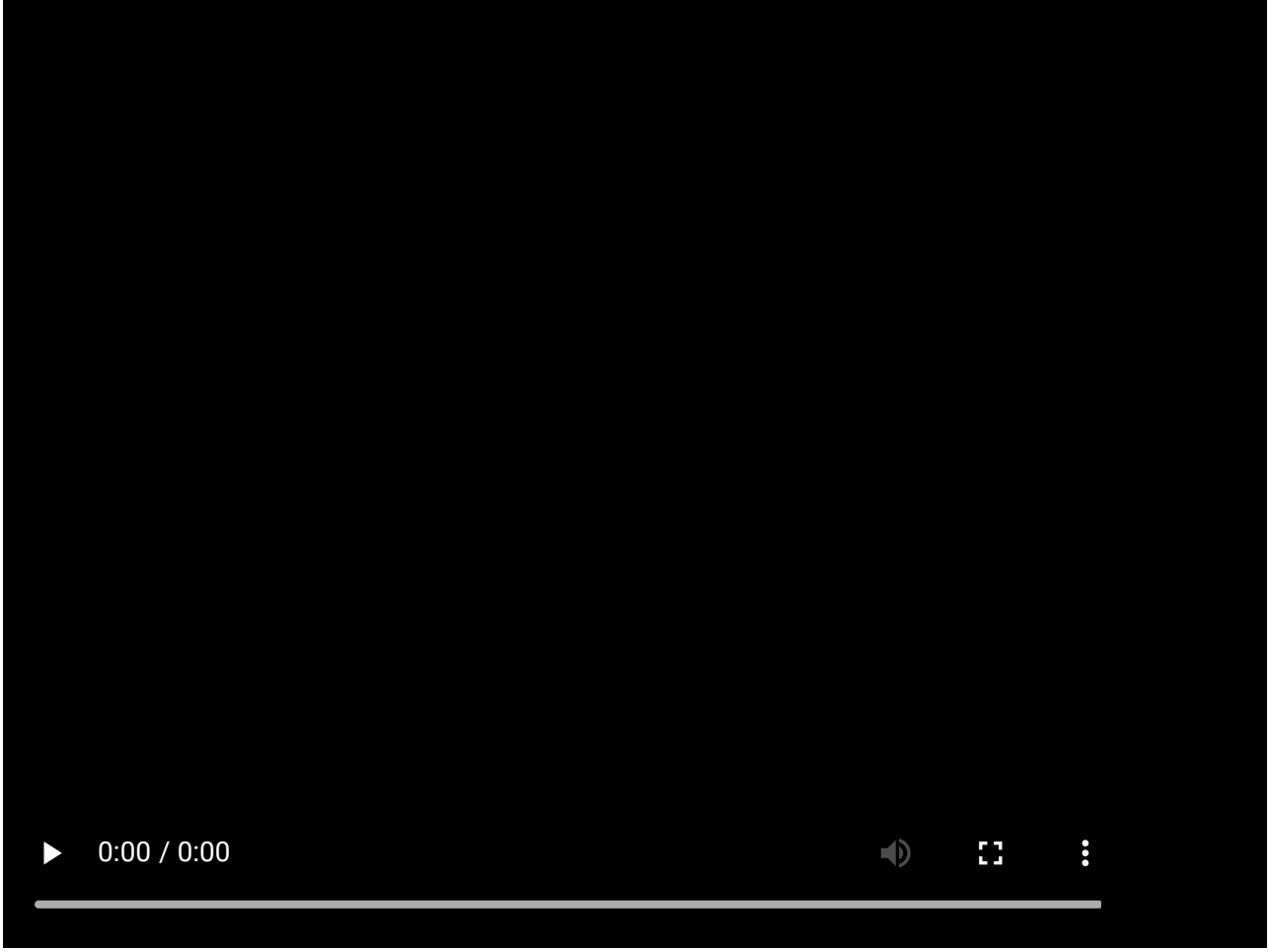

```
Episode 60,  score: -155.63566263169633
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

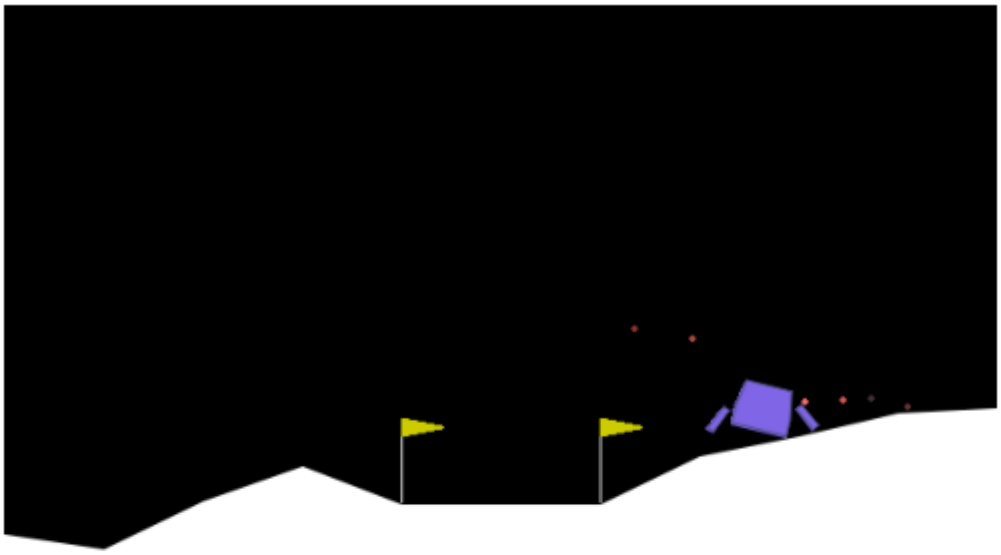


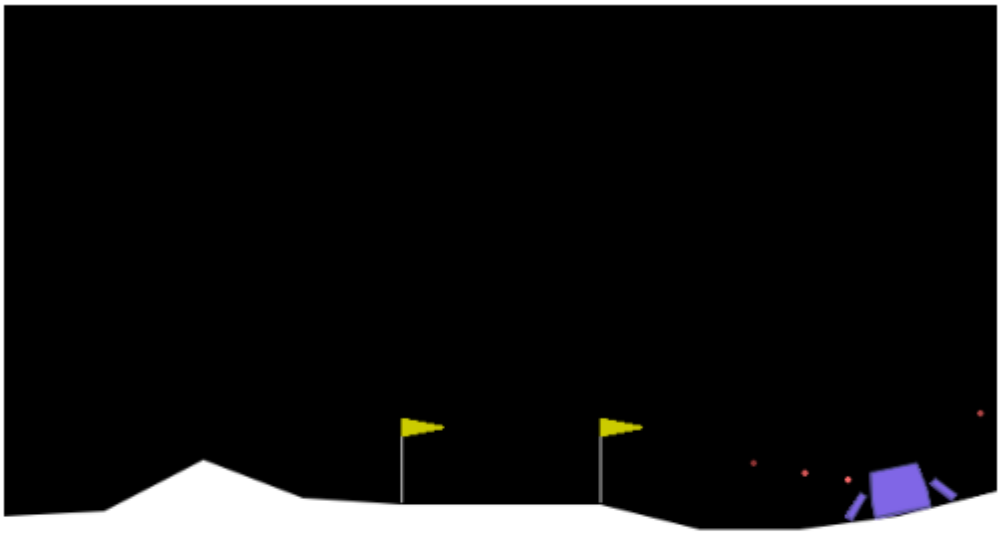

```
Episode 80,  score: -263.6509380219909
```

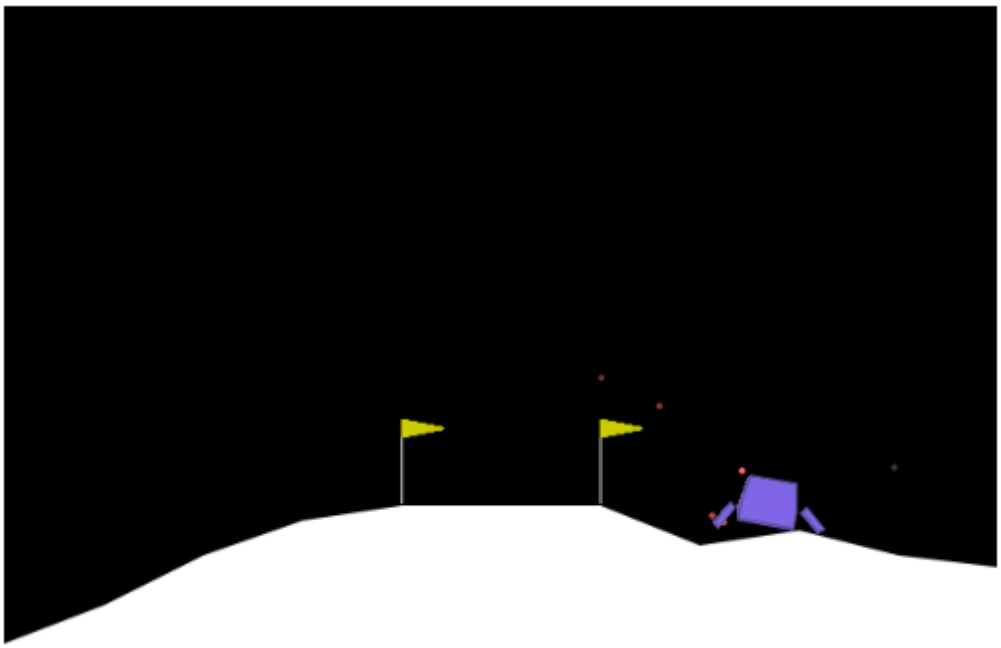
```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

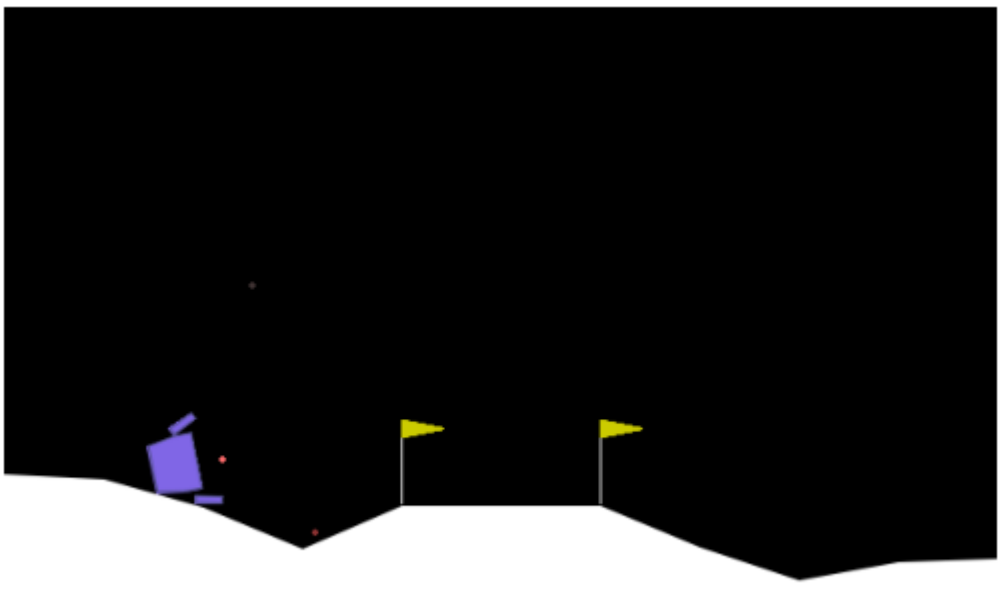


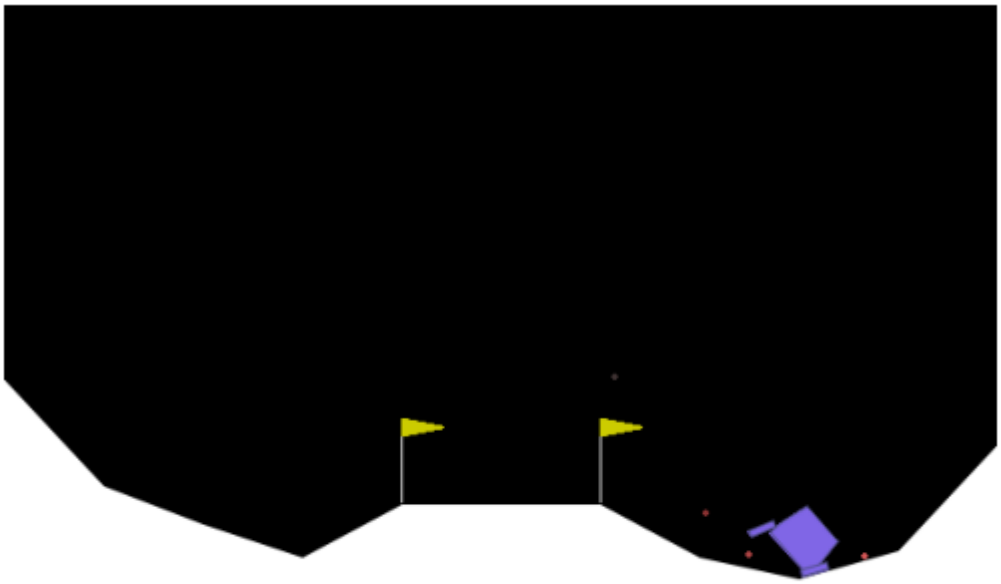
Episode 100, score: -217.36672609537007





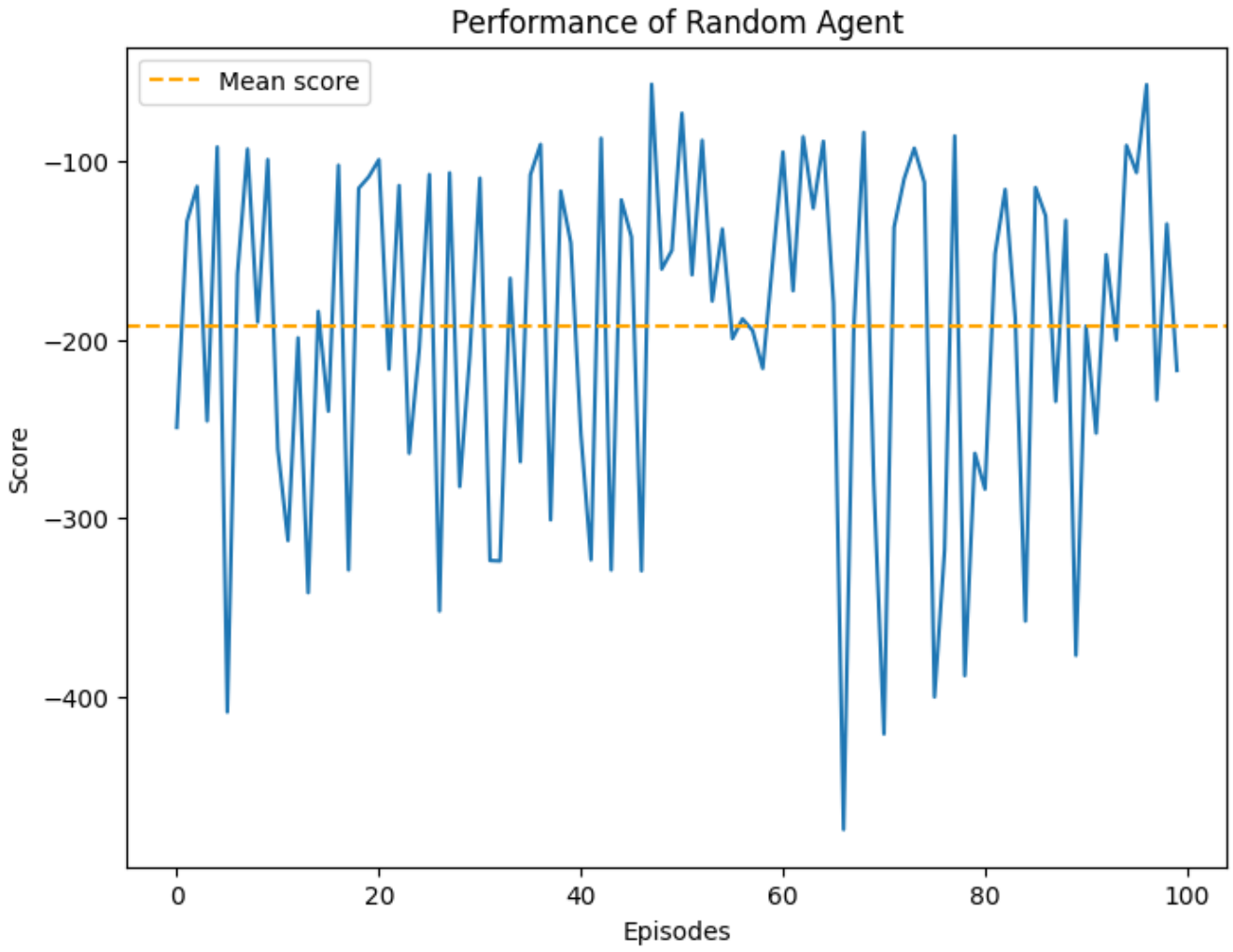






Visualising the performance of the agent

In [ ]:
```
plt.figure(figsize=(8,6))
plt.plot(range(max_episodes), scores)
plt.axhline(np.mean(scores), linestyle="--", color="orange", label="Mean score")
plt.title('Performance of Random Agent')
plt.xlabel('Episodes')
plt.ylabel('Score')
plt.legend()
plt.show()
print('Average score of random agent over {} episodes: {:.2f}'.format(max_episodes, np.mean(s
print(f"Best episode: {np.argmax(scores) + 1}\tScore: {np.max(scores)}")
```



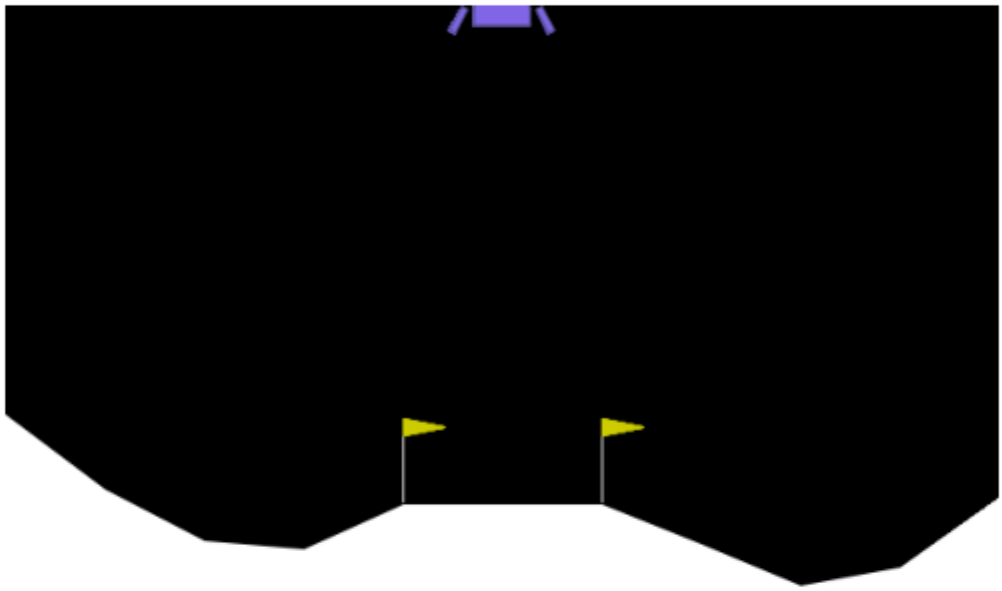
```
Average score of random agent over 100 episodes: -192.96
Best episode: 48        Score: -57.46388851591294
```

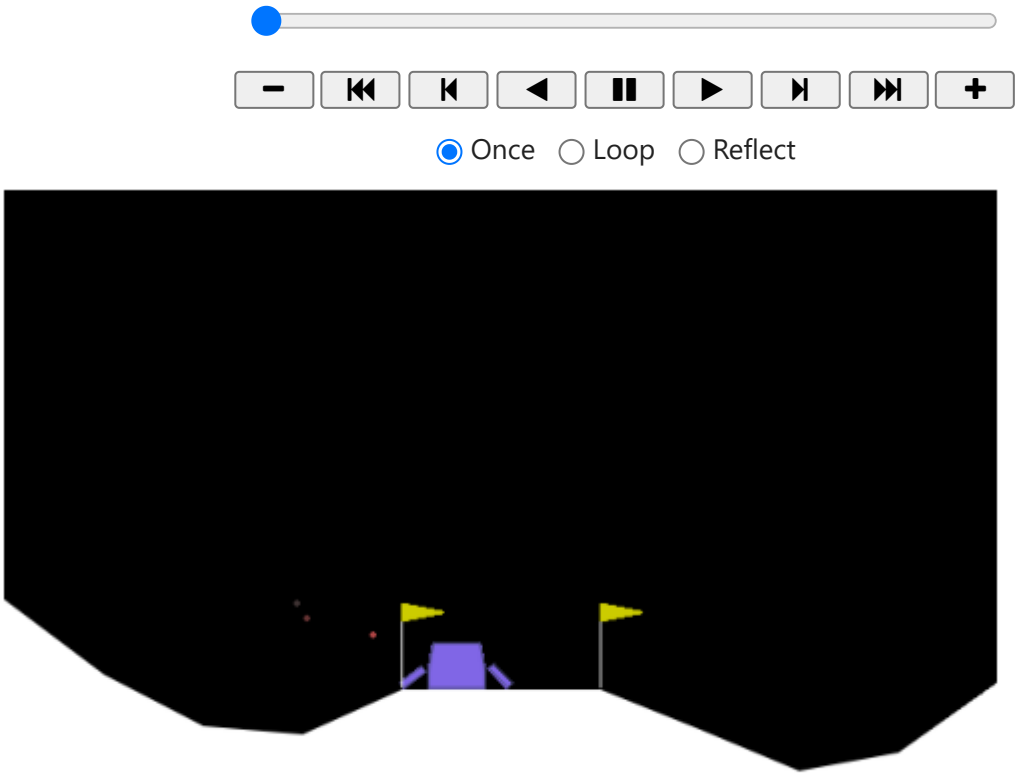
Observations

We note that as the baseline agent is an agent that performs random actions, we note that none of the models got above 0 as score. We also note that there is no clear learning that is happening as there is no AI model playing the game etc. This will be our baseline where the best episode is 48 and score is -57.46388851591294.

In [ ]:
```
create_animation(master_frames[np.argmax(scores)])
```

Out[ ]:





Observations

We can see that the lander score is reasonable as based on the scoring criteria the lander is close to the poles and has used very little engine fuel and as the lander crashed it reduced the score a lot.

Utility Functions

In [ ]:
```
BUFFER_SIZE = int(1e5)
BATCH_SIZE = 128
GAMMA = 0.99
TAU = 1e-3
LR = 5e-4
UPDATE_EVERY = 10
```

Replay Buffer

In [ ]:
```
class ReplayBuffer:

    def __init__(self, action_size, buffer_size, batch_size, seed):

        self.action_size = action_size
        self.memory = deque(maxlen=buffer_size)
        self.batch_size = batch_size
        self.experience = namedtuple("Experience", field_names=["state", "action", "reward",
        self.seed = random.seed(seed)

    def new_experience(self, state, action, reward, next_state, done):
        e = self.experience(state, action, reward, next_state, done)
        self.memory.append(e)

    def sample(self):

        experiences = random.sample(self.memory, k=self.batch_size)

        states = torch.from_numpy(np.vstack([e.state for e in experiences if e is not None]))
        actions = torch.from_numpy(np.vstack([e.action for e in experiences if e is not None]
        rewards = torch.from_numpy(np.vstack([e.reward for e in experiences if e is not None]
        next_states = torch.from_numpy(np.vstack([e.next_state for e in experiences if e is n
        dones = torch.from_numpy(np.vstack([e.done for e in experiences if e is not None]).as

        return (states, actions, rewards, next_states, dones)

    def __len__(self):

        return len(self.memory)
```

DQN Model

DQN, or Deep Q-Network, is a technique in reinforcement learning that uses both Q-Learning and deep neural networks. The network takes the current state of the environment as input and generates a prediction of the expected reward for each possible action. The objective is to find the best policy that leads to the maximum total reward. The algorithm uses a memory of past experiences, known as experience replay, and a separate network, called the target network, to maintain stability during the learning process. These techniques make DQN capable of handling complex and changing environments with high dimensions.

Create environment for the DQN model

In [ ]:
```
env = gym.make('LunarLander-v2',
               continuous = False,
               gravity = -10.0,
               enable_wind = False,
               wind_power = 15.0,
               turbulence_power = 1.5
)
env.seed(0)
```

Out[ ]:
```
[0]
```

Setup model architecture

In [ ]:
```
class QNetwork(nn.Module):

    def __init__(self, state_size, action_size, seed):
        """Initialize parameters and build model.
        Params
        ======
            state_size (int): Dimension of each state
            action_size (int): Dimension of each action
            seed (int): Random seed
        """
        super(QNetwork, self).__init__()
        self.seed = torch.manual_seed(seed)
        self.fc1 = nn.Linear(state_size, 128)
        self.fc2 = nn.Linear(128, 128)
        self.fc3 = nn.Linear(128, action_size)

    def forward(self, state):
        """Build a network that maps state -> action values."""
        x = self.fc1(state)
        x = F.relu(x)
        x = self.fc2(x)
        x = F.relu(x)
        return self.fc3(x)
```

In [ ]:
```
class DQNAgent():

    def __init__(self, state_size, action_size, seed):

        self.state_size = state_size
        self.action_size = action_size
        self.seed = random.seed(seed)

        self.qnetwork_local = QNetwork(state_size, action_size, seed).to(device)
        self.qnetwork_target = QNetwork(state_size, action_size, seed).to(device)
        self.optimizer = optim.Adam(self.qnetwork_local.parameters(), lr=LR)

        self.memory = ReplayBuffer(action_size, BUFFER_SIZE, BATCH_SIZE, seed)
        self.t_step = 0

    def step(self, state, action, reward, next_state, done):
        # Save experience in replay memory
        self.memory.new_experience(state, action, reward, next_state, done)

        # Learn every UPDATE_EVERY time steps.
        self.t_step = (self.t_step + 1) % UPDATE_EVERY
        if self.t_step == 0:
            # If enough samples are available in memory, get random subset and learn
            if len(self.memory) > BATCH_SIZE:
                experiences = self.memory.sample()
                self.learn(experiences, GAMMA)

    def act(self, state, eps=0.):

        state = torch.from_numpy(state).float().unsqueeze(0).to(device)
        self.qnetwork_local.eval()
        with torch.no_grad():
            action_values = self.qnetwork_local(state)
        self.qnetwork_local.train()

        # Epsilon-greedy action selection
        if random.random() > eps:
            return np.argmax(action_values.cpu().data.numpy())
        else:
            return random.choice(np.arange(self.action_size))
```

```
    def learn(self, experiences, gamma):

        # Obtain random minibatch of tuples from D
        states, actions, rewards, next_states, dones = experiences

        ## Compute and minimize the loss
        q_targets_next = self.qnetwork_target(next_states).detach().max(1)[0].unsqueeze(1)
        q_targets = rewards + gamma * q_targets_next * (1 - dones)
        q_expected = self.qnetwork_local(states).gather(1, actions)

        ### Loss calculation (we used Mean squared error)
        loss = F.mse_loss(q_expected, q_targets)
        self.optimizer.zero_grad()
        loss.backward()
        self.optimizer.step()

        # ------------------- update target network ------------------- #
        self.soft_update(self.qnetwork_local, self.qnetwork_target, TAU)

    def soft_update(self, local_model, target_model, tau):

        for target_param, local_param in zip(target_model.parameters(), local_model.parameter
            target_param.data.copy_(tau*local_param.data + (1.0-tau)*target_param.data)
```

In [ ]:

```
class ReplayBuffer:

    def __init__(self, action_size, buffer_size, batch_size, seed):

        self.action_size = action_size
        self.memory = deque(maxlen=buffer_size)
        self.batch_size = batch_size
        self.experience = namedtuple("Experience", field_names=["state", "action", "reward",
        self.seed = random.seed(seed)

    def new_experience(self, state, action, reward, next_state, done):
        e = self.experience(state, action, reward, next_state, done)
        self.memory.append(e)

    def sample(self):

        experiences = random.sample(self.memory, k=self.batch_size)

        states = torch.from_numpy(np.vstack([e.state for e in experiences if e is not None]))
        actions = torch.from_numpy(np.vstack([e.action for e in experiences if e is not None]
        rewards = torch.from_numpy(np.vstack([e.reward for e in experiences if e is not None]
        next_states = torch.from_numpy(np.vstack([e.next_state for e in experiences if e is n
        dones = torch.from_numpy(np.vstack([e.done for e in experiences if e is not None]).as

        return (states, actions, rewards, next_states, dones)

    def __len__(self):

        return len(self.memory)
```

In [ ]:

```
def train(n_episodes=2000, max_t=1000, eps_start=1.0, eps_end=0.01, eps_decay=0.995):

    master_frames = []
    scores = []
    scores_window = deque(maxlen=100)
    eps = eps_start
    for i_episode in range(1, n_episodes+1):
        state = env.reset()
        score = 0
        if i_episode % 100 == 0:
            frames = []
        for t in range(max_t):
            action = agent.act(state, eps)
            next_state, reward, done, _ = env.step(action)
```

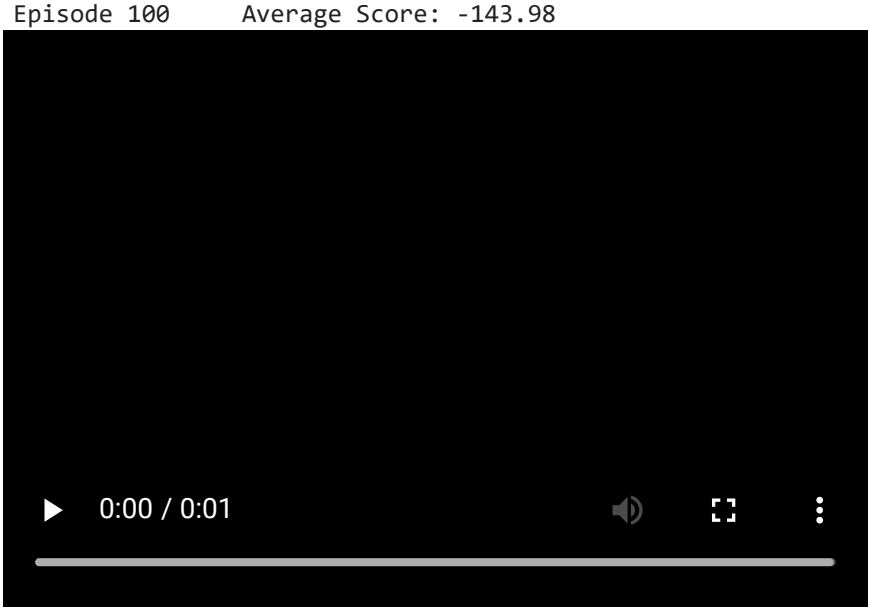
```
            agent.step(state, action, reward, next_state, done)
            if i_episode % 100 == 0:
                screen = env.render(mode='rgb_array')
                frames.append(screen)
            state = next_state
            score += reward
            if done:
                break
        scores_window.append(score)       # save most recent score
        scores.append(score)              # save most recent score
        eps = max(eps_end, eps_decay*eps) # decrease epsilon
        print('\rEpisode {}\tAverage Score: {:.2f}'.format(i_episode, np.mean(scores_window)))
        if i_episode % 100 == 0:
            create_animation(frames)
            master_frames.append(frames)
            print('\rEpisode {}\tAverage Score: {:.2f}'.format(i_episode, np.mean(scores_wind
        torch.save(agent.qnetwork_local.state_dict(), 'checkpoint.pth')
    return scores, master_frames
```

In [ ]:
```
agent = DQNAgent(state_size=8, action_size=4, seed=0)
scores, master_frames = train()
```

```
Episode 99      Average Score: -144.15
```

```
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

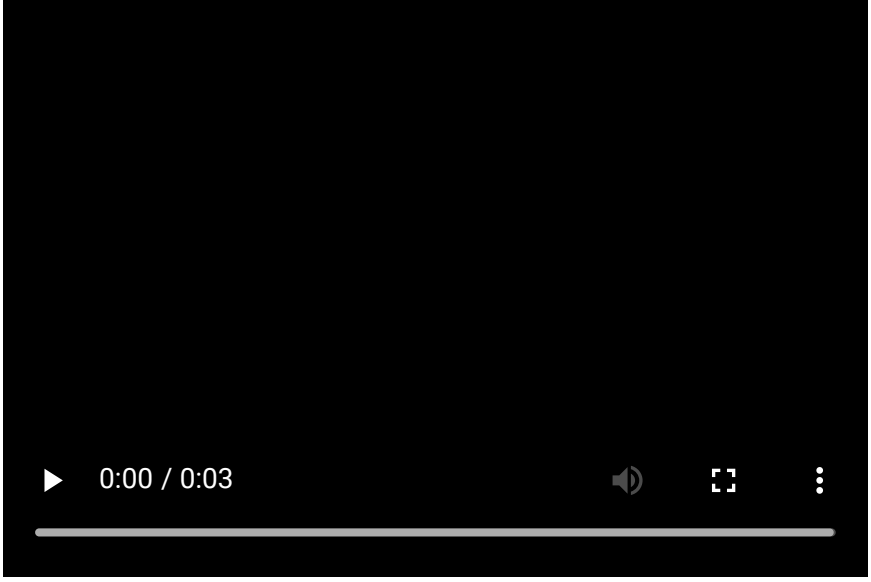
```
Episode 100     Average Score: -143.98
```



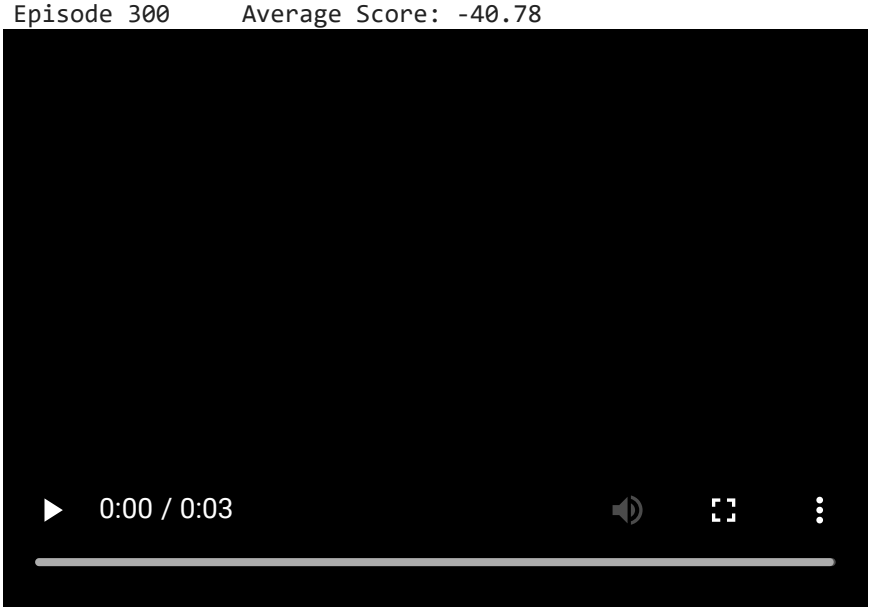

```
Episode 100     Average Score: -143.98
Episode 199     Average Score: -113.78
```

```
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

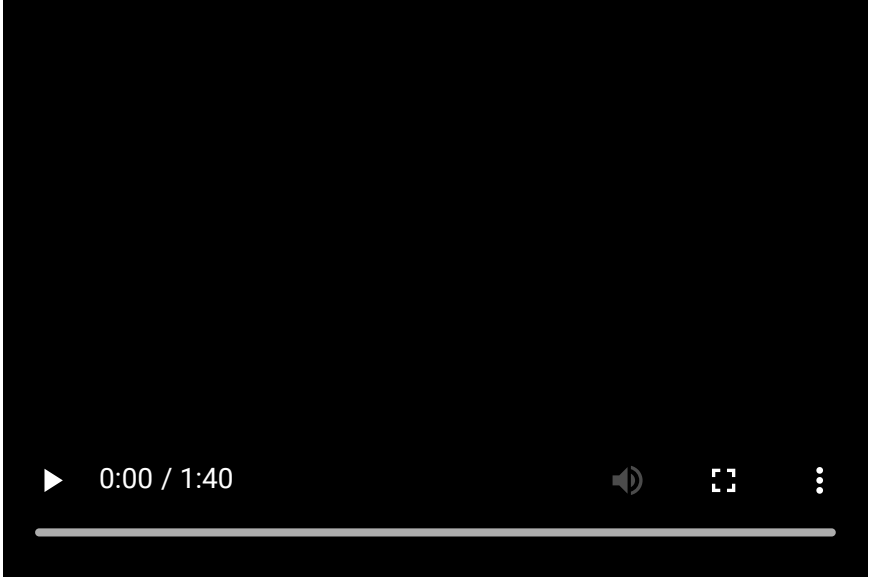
```
Episode 200     Average Score: -113.15
```



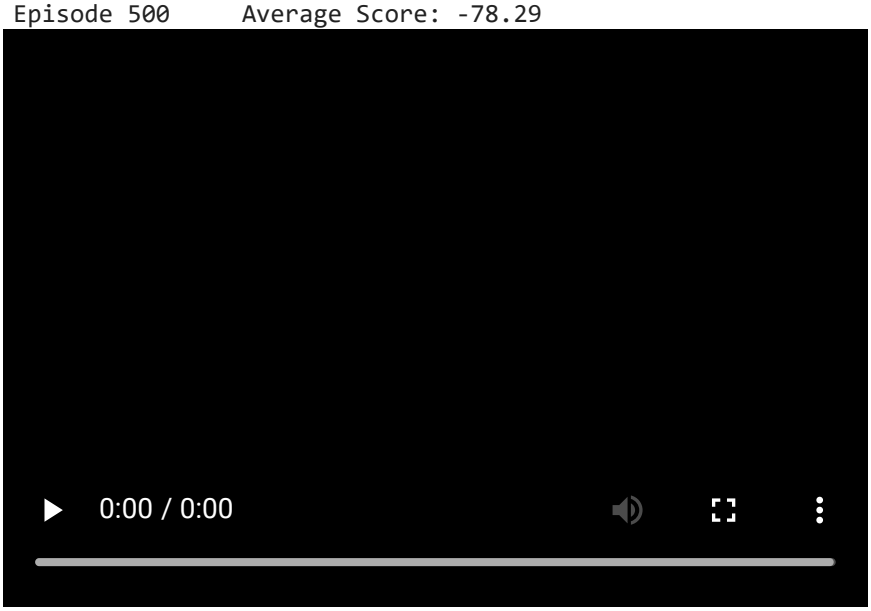

```
Episode 200     Average Score: -113.15
Episode 299     Average Score: -39.660
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 300     Average Score: -40.78
```



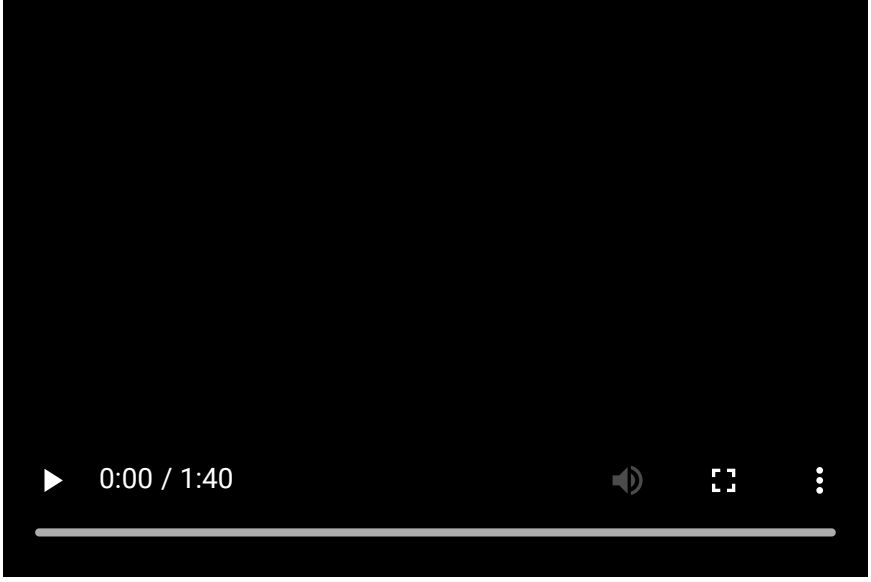

```
Episode 300     Average Score: -40.78
Episode 399     Average Score: -74.57
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 400     Average Score: -74.10
```



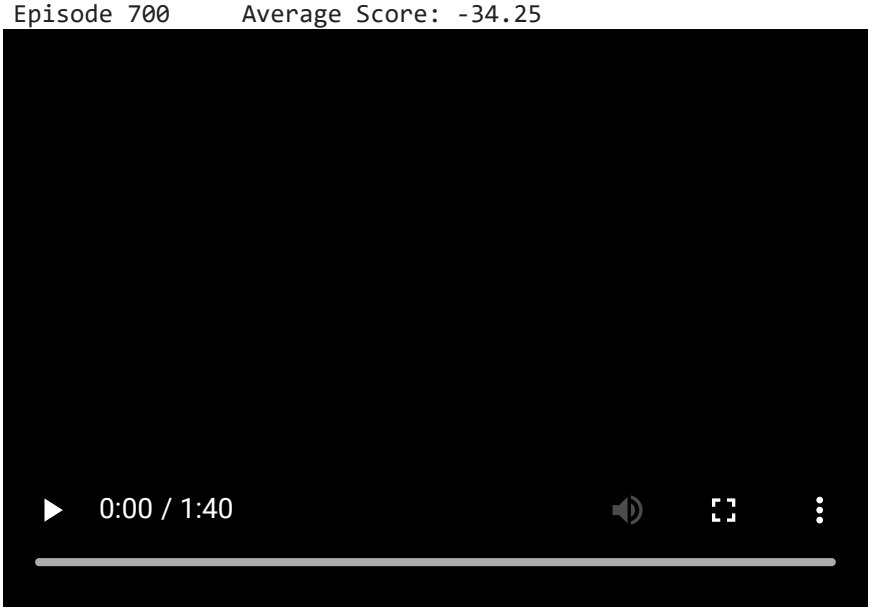

```
Episode 400     Average Score: -74.10
Episode 499     Average Score: -78.36
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 500     Average Score: -78.29
```



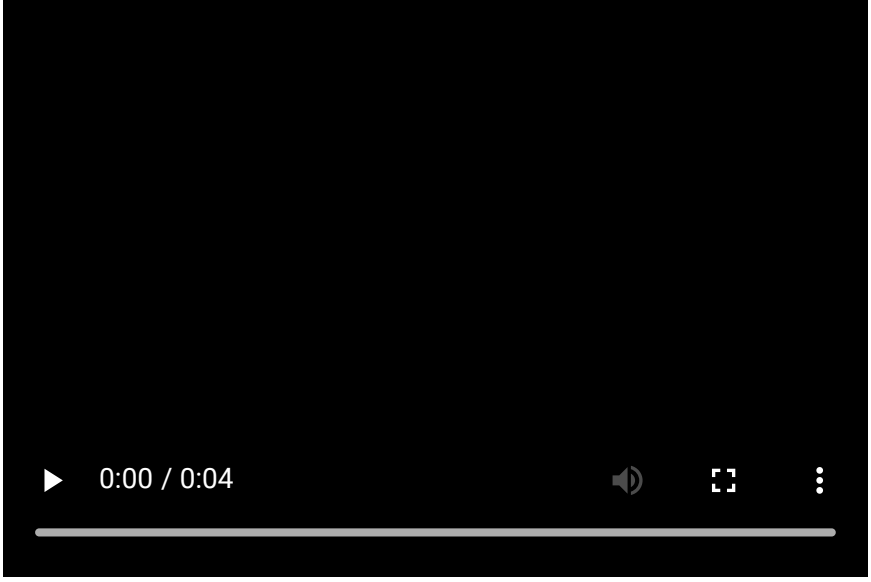

```
Episode 500     Average Score: -78.29
Episode 599     Average Score: -62.64
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 600     Average Score: -61.65
```



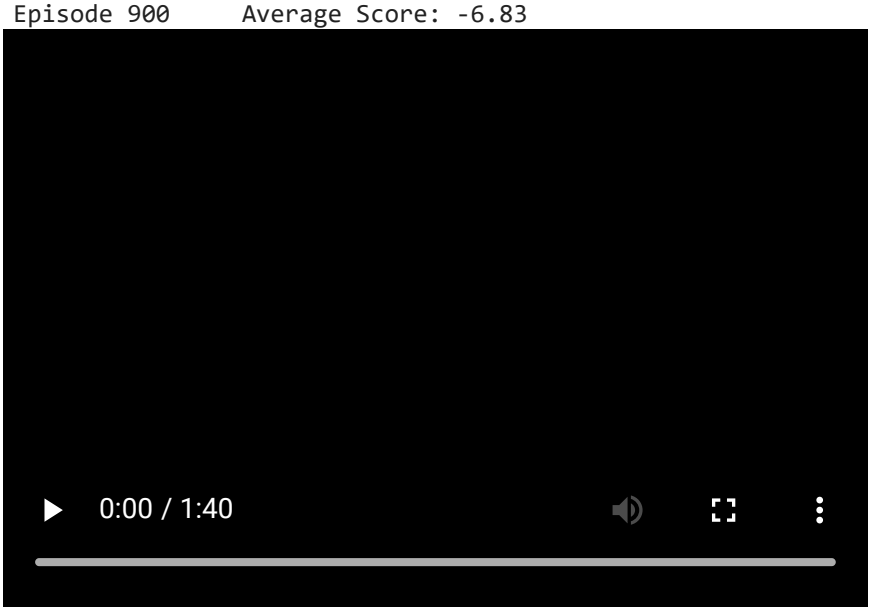

```
Episode 600     Average Score: -61.65
Episode 699     Average Score: -34.14
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 700     Average Score: -34.25
```



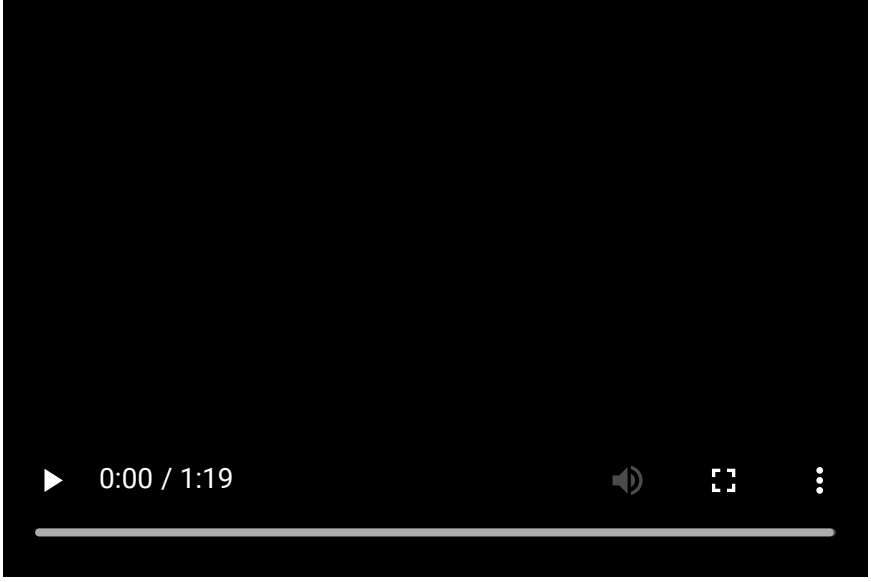

```
Episode 700     Average Score: -34.25
Episode 799     Average Score: -5.204
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 800     Average Score: -5.49
```



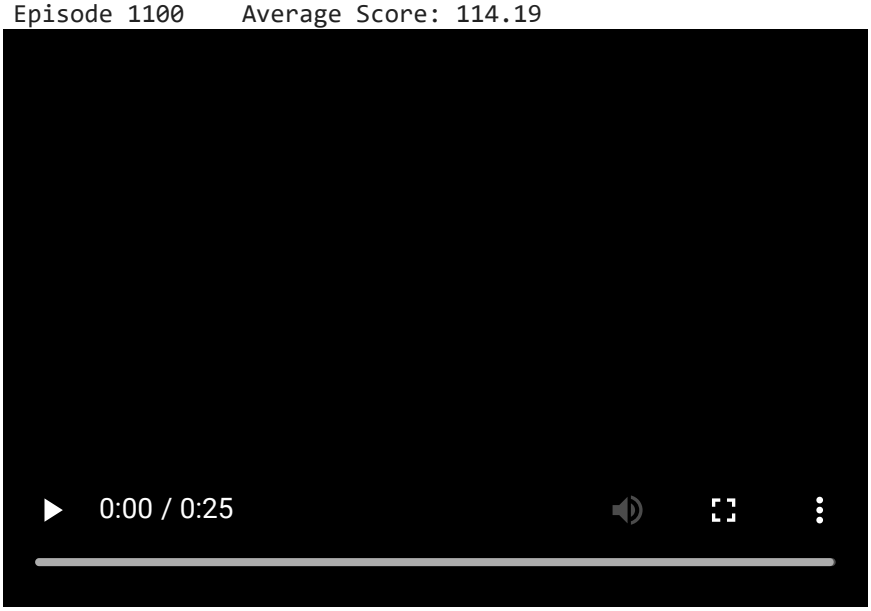

```
Episode 800     Average Score: -5.49
Episode 899     Average Score: -7.06
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 900     Average Score: -6.83
```



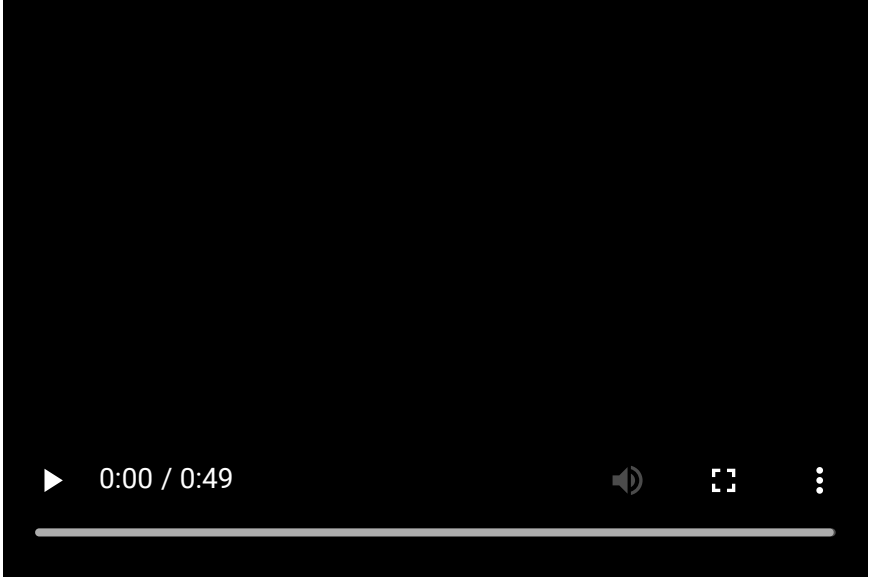

```
Episode 900     Average Score: -6.83
Episode 999     Average Score: 4.5029
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1000    Average Score: 6.02
```



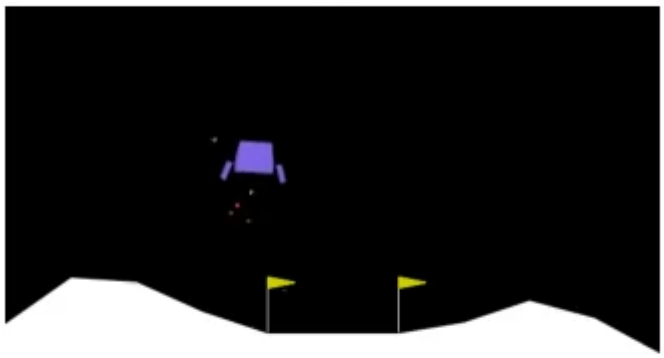

```
Episode 1000    Average Score: 6.02
Episode 1099    Average Score: 113.63
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1100    Average Score: 114.19
```



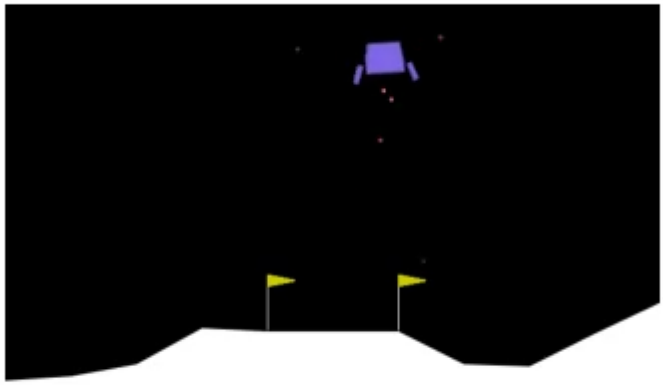

```
Episode 1100    Average Score: 114.19
Episode 1199    Average Score: 128.48
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1200    Average Score: 128.66
```



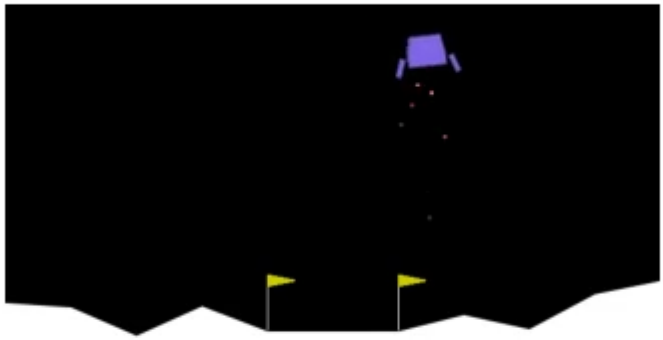

```
Episode 1200    Average Score: 128.66
Episode 1299    Average Score: 130.06
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1300    Average Score: 129.97
```



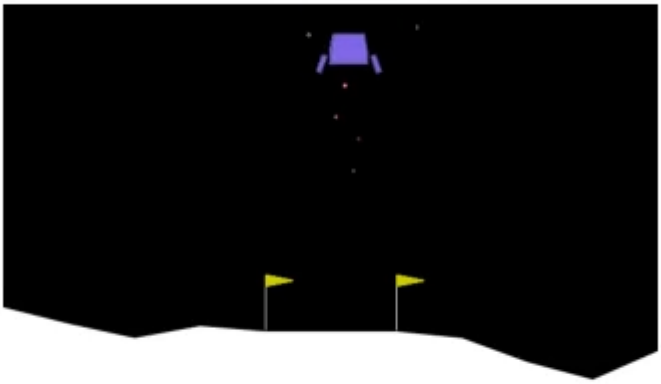
```
Episode 1300    Average Score: 129.97
Episode 1399    Average Score: 137.76
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1400    Average Score: 135.62
```



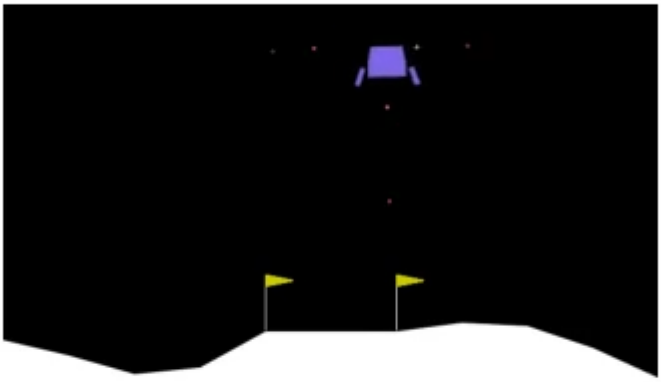
```
Episode 1400    Average Score: 135.62
Episode 1499    Average Score: 147.23
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1500    Average Score: 147.21
```



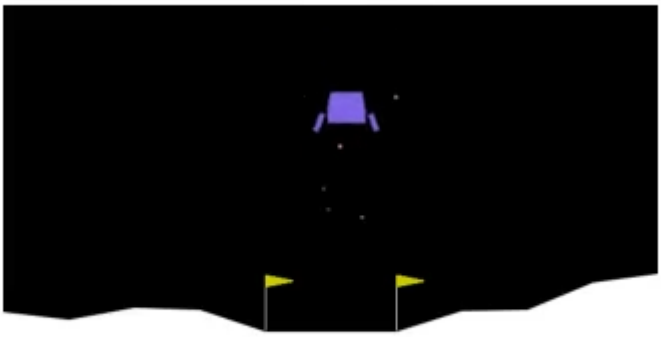
```
Episode 1500    Average Score: 147.21
Episode 1599    Average Score: 129.99
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1600    Average Score: 132.15
```



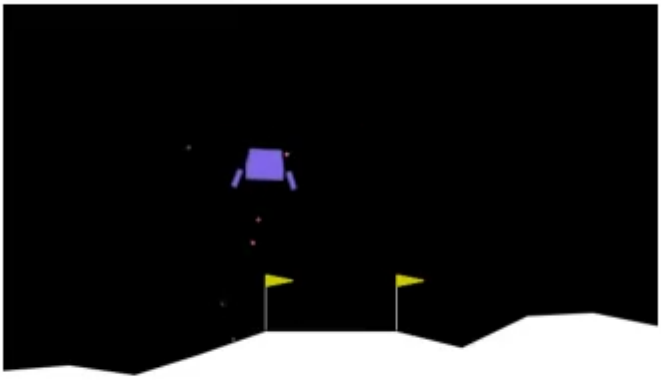
```
Episode 1600    Average Score: 132.15
Episode 1699    Average Score: 124.94
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1700    Average Score: 122.89
```



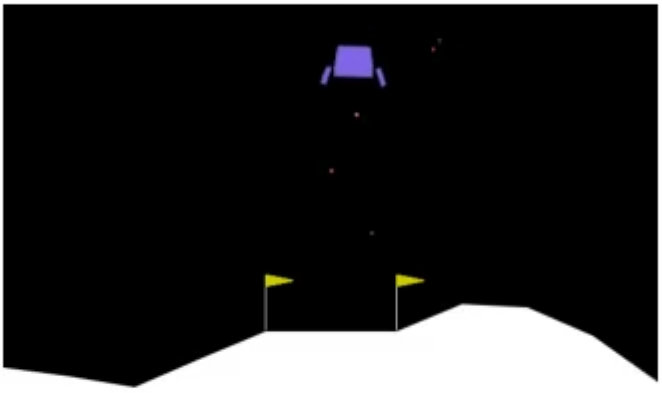
```
Episode 1700    Average Score: 122.89
Episode 1799    Average Score: 142.17
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1800    Average Score: 144.23
```



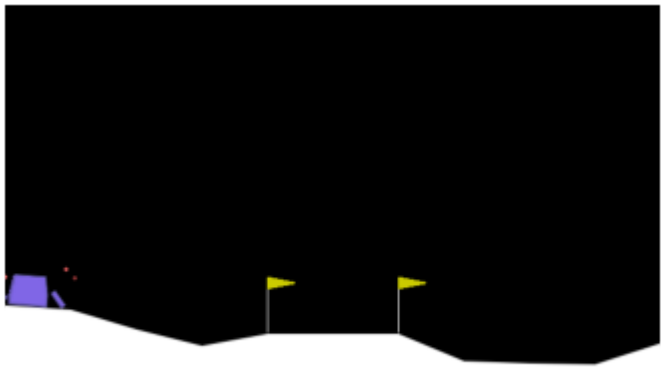
```
Episode 1800    Average Score: 144.23
Episode 1899    Average Score: 129.83
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1900    Average Score: 130.23
```

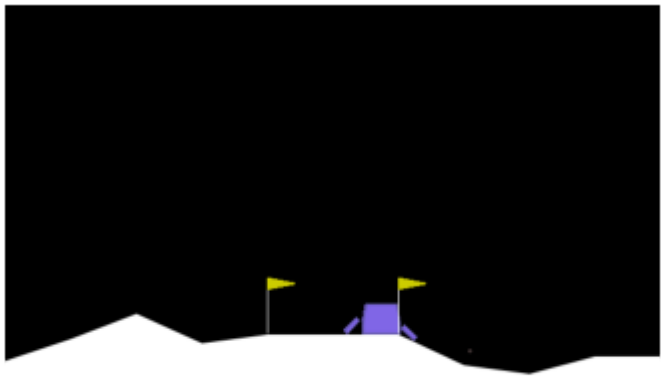


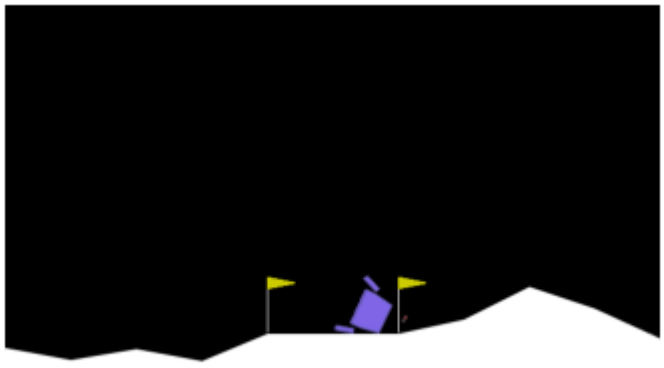
```
Episode 1900    Average Score: 130.23
Episode 1999    Average Score: 173.63
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2000    Average Score: 173.30
```

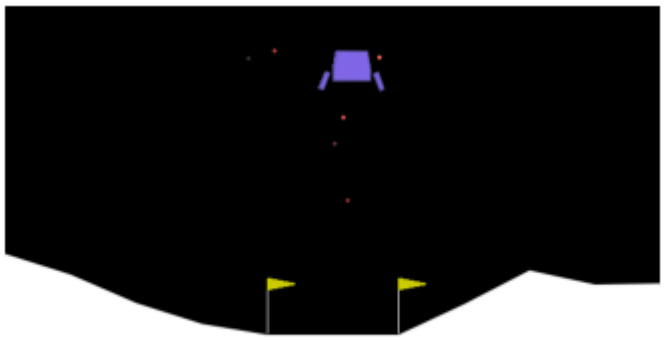


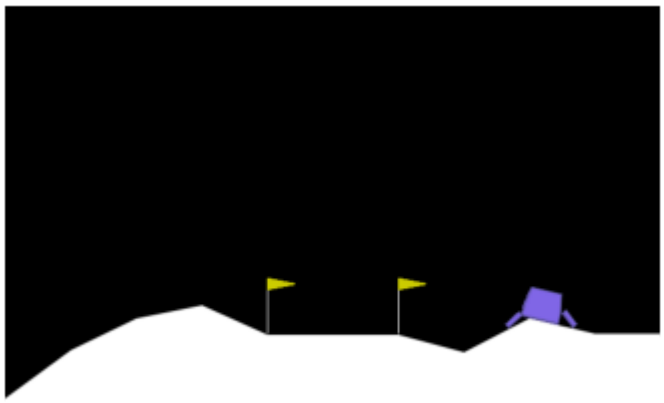
Episode 2000 Average Score: 173.30

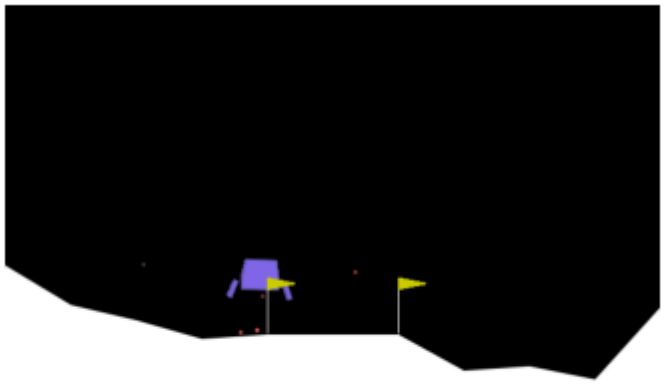


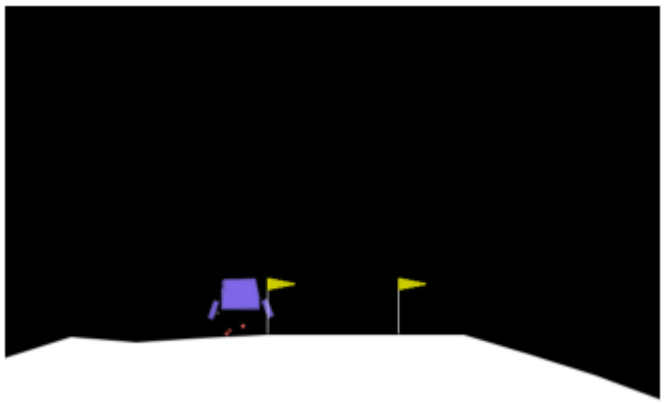


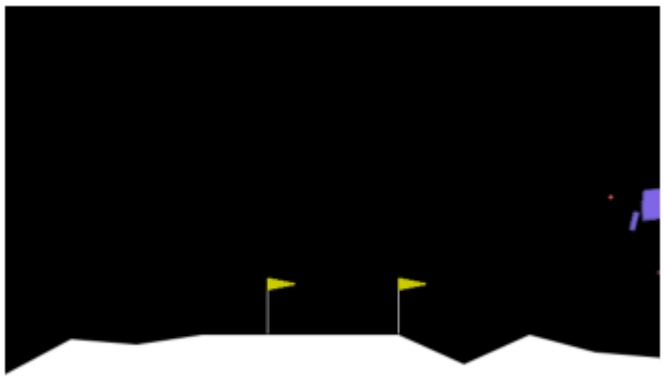


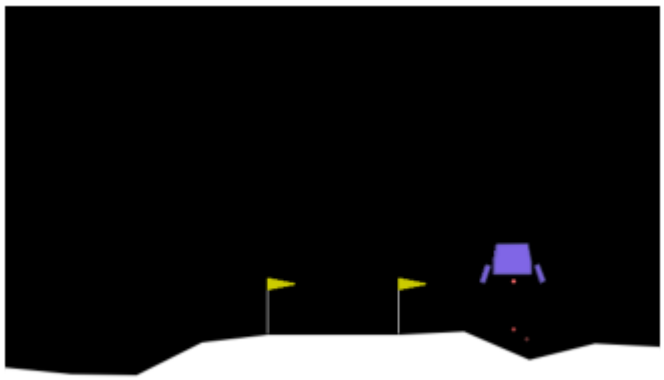


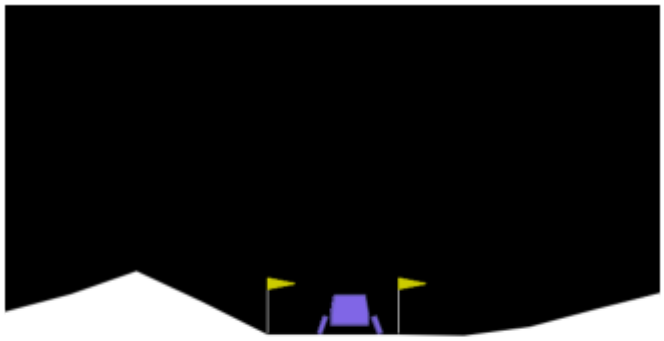


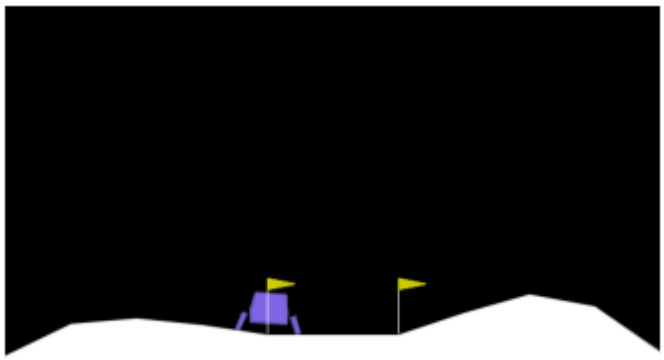


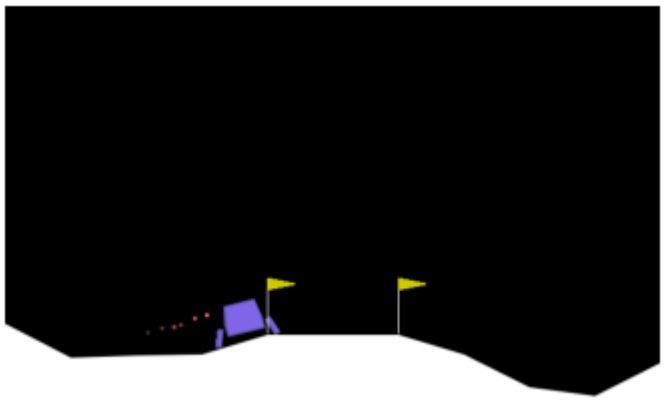


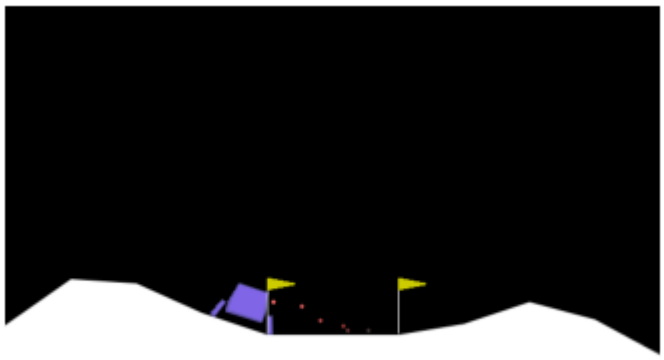


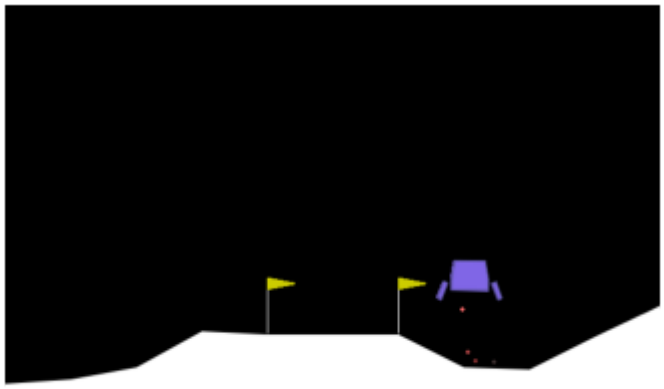


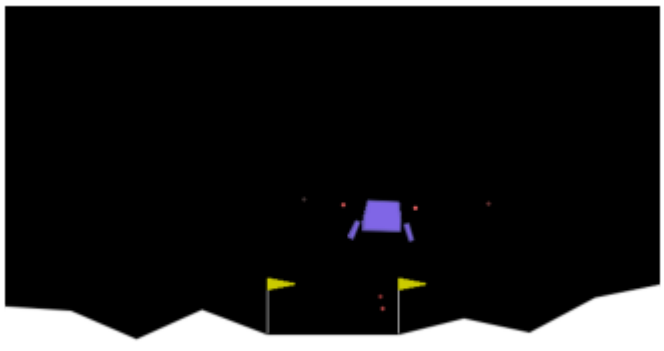


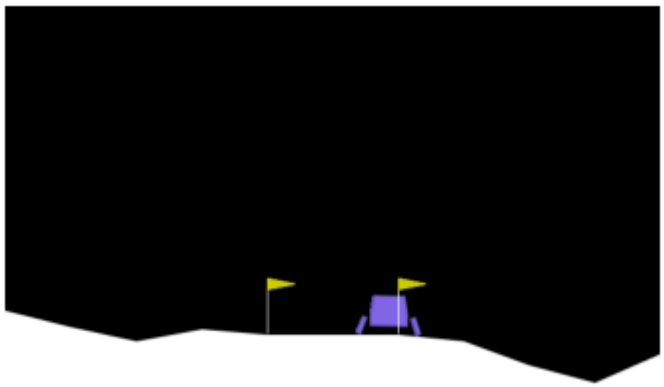


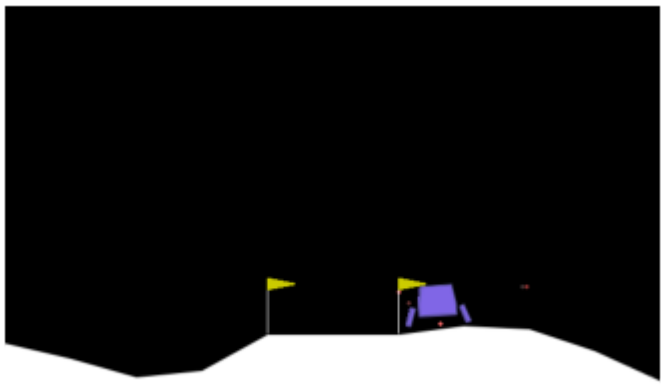


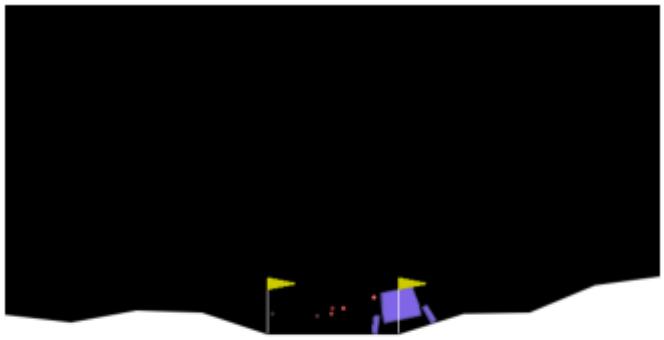


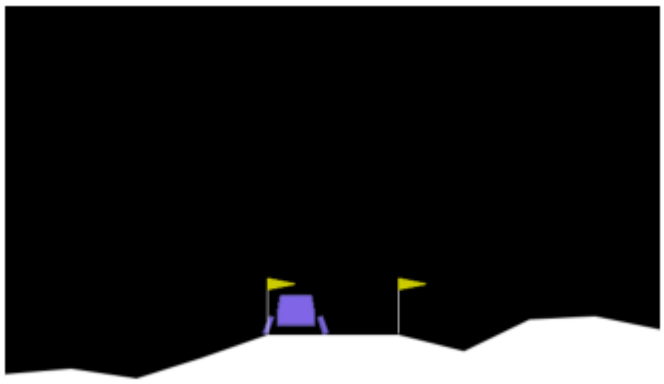


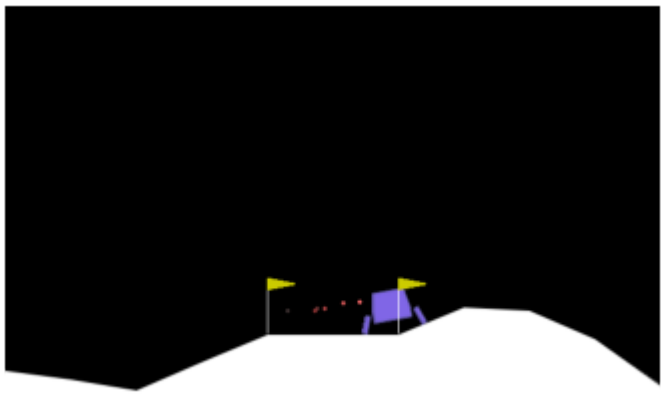






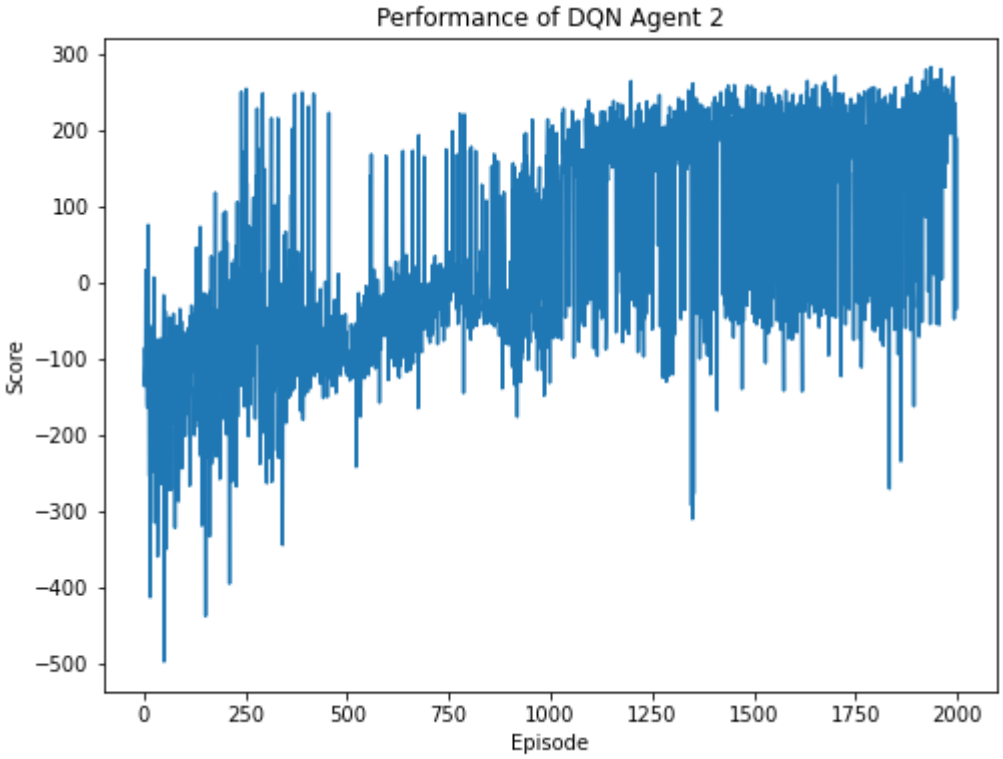






In [ ]:
```
fig = plt.figure(figsize=(8,6))
ax = fig.add_subplot(111)
plt.plot(np.arange(len(scores)), scores)
plt.ylabel('Score')
plt.xlabel('Episode')
plt.title('Performance of DQN Agent')
plt.show()
```

```
/opt/conda/lib/python3.8/site-packages/ipykernel/ipkernel.py:287: DeprecationWarning: `should
_run_async` will not call `transform_cell` automatically in the future. Please pass the resul
t to `transformed_cell` argument and any exception that happen during thetransform in `prepro
cessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```



Improvement

Here we added an extra hidden layer.

In [ ]:
```
class QNetwork_2(nn.Module):

    def __init__(self, state_size, action_size, seed):
        """Initialize parameters and build model.
        Params
        ======
            state_size (int): Dimension of each state
            action_size (int): Dimension of each action
            seed (int): Random seed
```

```
        """
        super(QNetwork_2, self).__init__()
        self.seed = torch.manual_seed(seed)
        self.fc1 = nn.Linear(state_size, 128)
        self.fc2 = nn.Linear(128, 128)
        self.fc3 = nn.Linear(128, 128)
        self.fc4 = nn.Linear(128, action_size)

    def forward(self, state):
        """Build a network that maps state -> action values."""
        x = self.fc1(state)
        x = F.relu(x)
        x = self.fc2(x)
        x = F.relu(x)
        x = self.fc3(x)
        x = F.relu(x)
        return self.fc4(x)
```

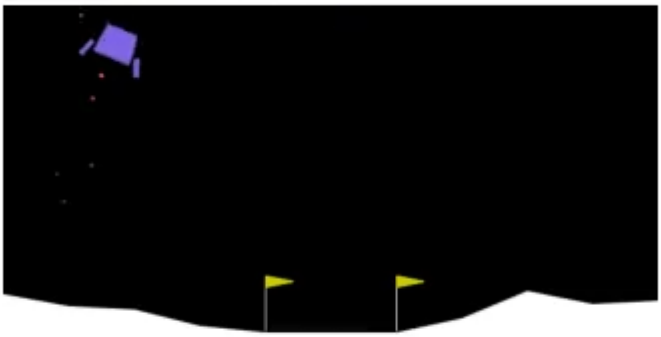
In [ ]:

```
class DQNAgent_2():

    def __init__(self, state_size, action_size, seed):

        self.state_size = state_size
        self.action_size = action_size
        self.seed = random.seed(seed)

        self.qnetwork_local = QNetwork_2(state_size, action_size, seed).to(device)
        self.qnetwork_target = QNetwork_2(state_size, action_size, seed).to(device)
        self.optimizer = optim.Adam(self.qnetwork_local.parameters(), lr=LR)

        self.memory = ReplayBuffer(action_size, BUFFER_SIZE, BATCH_SIZE, seed)
        self.t_step = 0

    def step(self, state, action, reward, next_state, done):
        # Save experience in replay memory
        self.memory.new_experience(state, action, reward, next_state, done)

        # Learn every UPDATE_EVERY time steps.
        self.t_step = (self.t_step + 1) % UPDATE_EVERY
        if self.t_step == 0:
            # If enough samples are available in memory, get random subset and learn
            if len(self.memory) > BATCH_SIZE:
                experiences = self.memory.sample()
                self.learn(experiences, GAMMA)

    def act(self, state, eps=0.):

        state = torch.from_numpy(state).float().unsqueeze(0).to(device)
        self.qnetwork_local.eval()
        with torch.no_grad():
            action_values = self.qnetwork_local(state)
        self.qnetwork_local.train()

        # Epsilon-greedy action selection
        if random.random() > eps:
            return np.argmax(action_values.cpu().data.numpy())
        else:
            return random.choice(np.arange(self.action_size))

    def learn(self, experiences, gamma):

        # Obtain random minibatch of tuples from D
        states, actions, rewards, next_states, dones = experiences

        ## Compute and minimize the loss
        q_targets_next = self.qnetwork_target(next_states).detach().max(1)[0].unsqueeze(1)
        q_targets = rewards + gamma * q_targets_next * (1 - dones)
        q_expected = self.qnetwork_local(states).gather(1, actions)
```

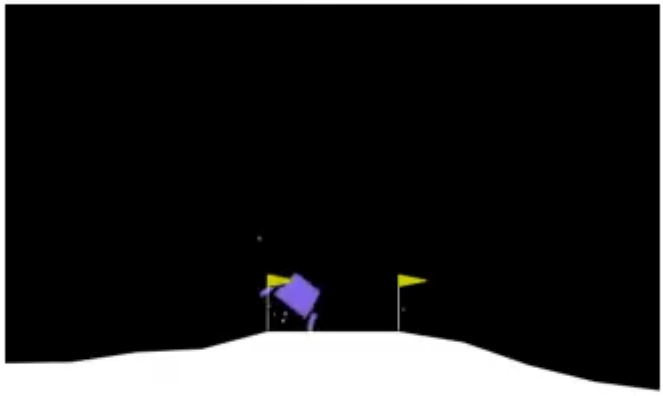
```
        ### Loss calculation (we used Mean squared error)
        loss = F.mse_loss(q_expected, q_targets)
        self.optimizer.zero_grad()
        loss.backward()
        self.optimizer.step()

        # ------------------- update target network ------------------- #
        self.soft_update(self.qnetwork_local, self.qnetwork_target, TAU)

    def soft_update(self, local_model, target_model, tau):

        for target_param, local_param in zip(target_model.parameters(), local_model.parameter
            target_param.data.copy_(tau*local_param.data + (1.0-tau)*target_param.data)
```

In [ ]:
```
agent = DQNAgent_2(state_size=8, action_size=4, seed=0)
scores, master_frames = train()
```

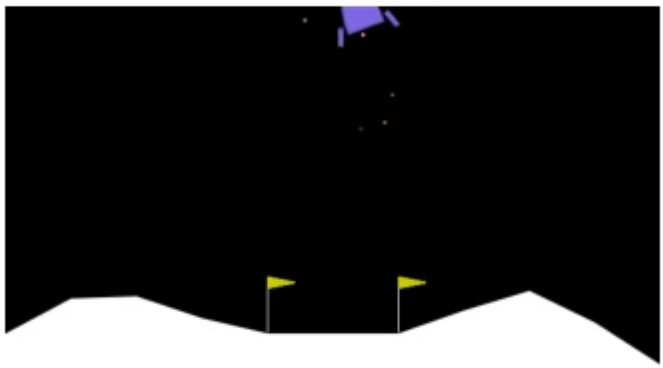
```
/opt/conda/lib/python3.8/site-packages/gym/utils/passive_env_checker.py:241: DeprecationWarni
ng: `np.bool8` is a deprecated alias for `np.bool_`.  (Deprecated NumPy 1.24)
  if not isinstance(terminated, (bool, np.bool8)):
Episode 99      Average Score: -176.59
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument
mode in render method is deprecated; use render_mode during environment initialization instea
d.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 100     Average Score: -177.77
```



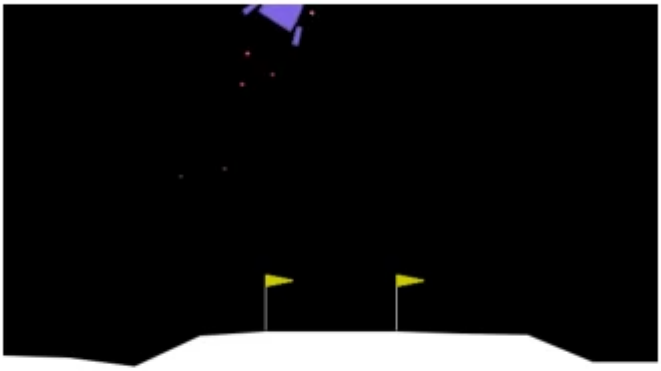
```
Episode 100     Average Score: -177.77
Episode 199     Average Score: -124.39
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument
mode in render method is deprecated; use render_mode during environment initialization instea
d.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 200     Average Score: -121.59
```



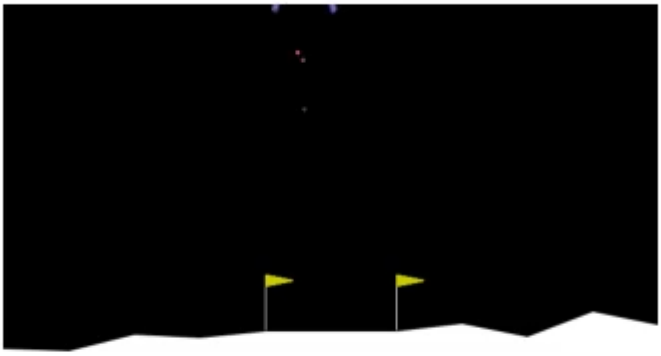
```
Episode 200     Average Score: -121.59
Episode 299     Average Score: -95.189
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 300     Average Score: -94.25
```



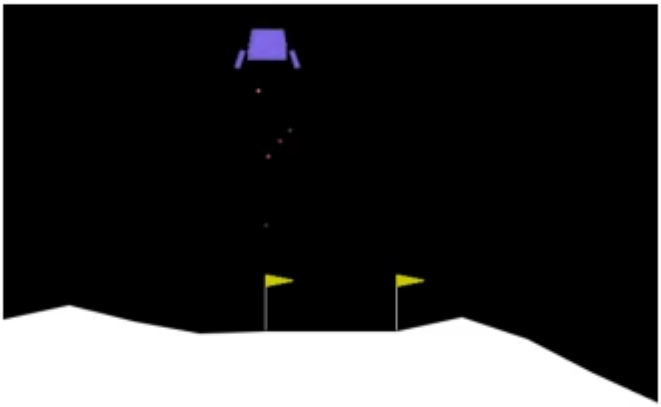
```
Episode 300     Average Score: -94.25
Episode 399     Average Score: -42.38
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 400     Average Score: -43.52
```



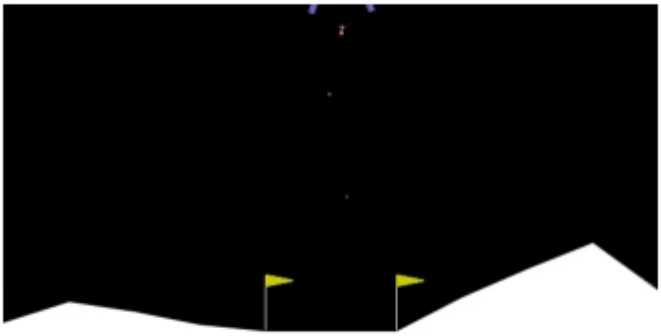
```
Episode 400     Average Score: -43.52
Episode 499     Average Score: -55.04
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 500     Average Score: -55.01
```



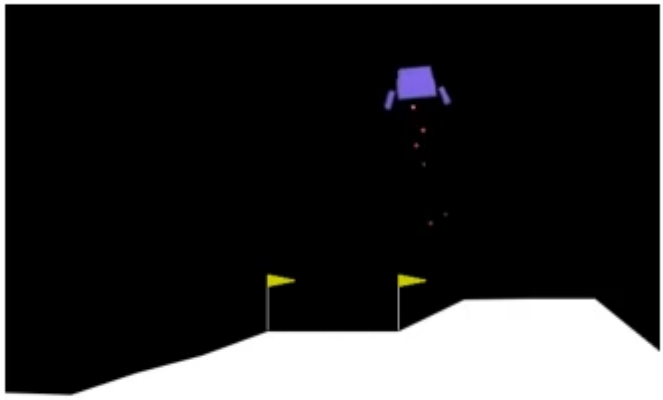
```
Episode 500     Average Score: -55.01
Episode 599     Average Score: -58.07
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 600     Average Score: -57.79
```



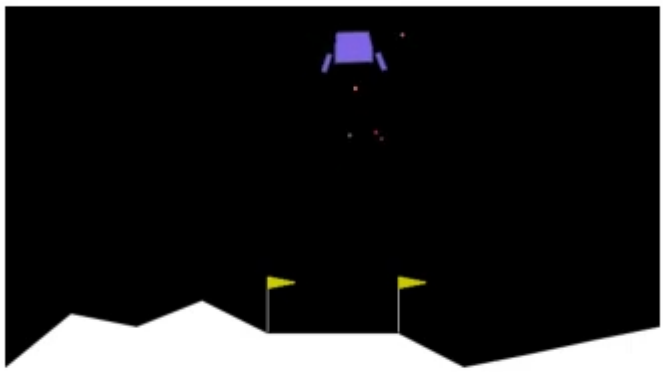
```
Episode 600     Average Score: -57.79
Episode 699     Average Score: -46.67
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 700     Average Score: -46.85
```



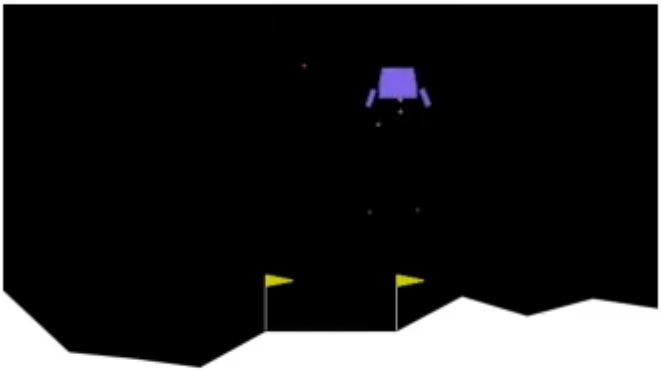
```
Episode 700     Average Score: -46.85
Episode 799     Average Score: -4.425
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 800     Average Score: -4.08
```



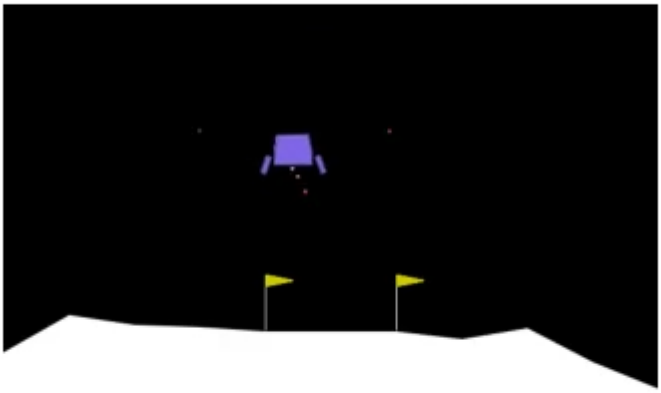
```
Episode 800     Average Score: -4.08
Episode 899     Average Score: 19.36
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 900     Average Score: 18.47
```



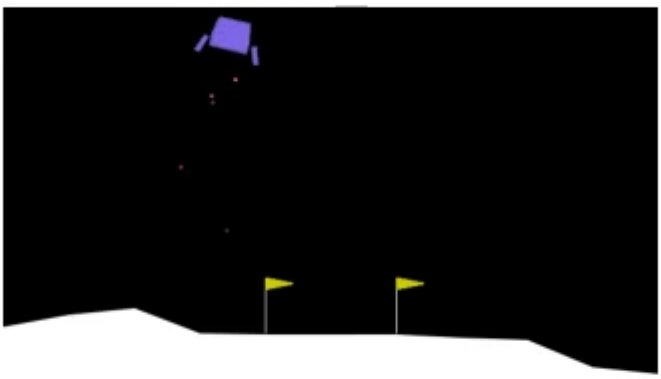
```
Episode 900     Average Score: 18.47
Episode 999     Average Score: 45.45
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1000    Average Score: 47.31
```



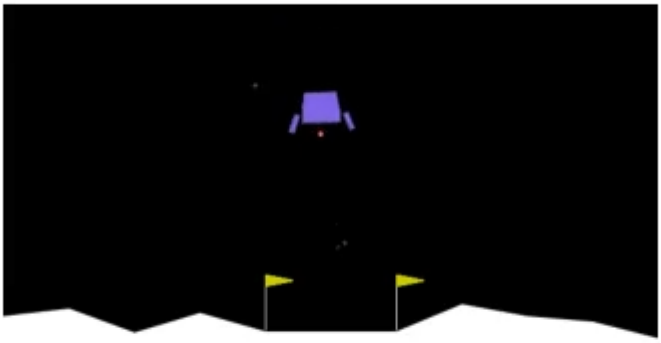
```
Episode 1000    Average Score: 47.31
Episode 1099    Average Score: 111.06
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1100    Average Score: 111.85
```



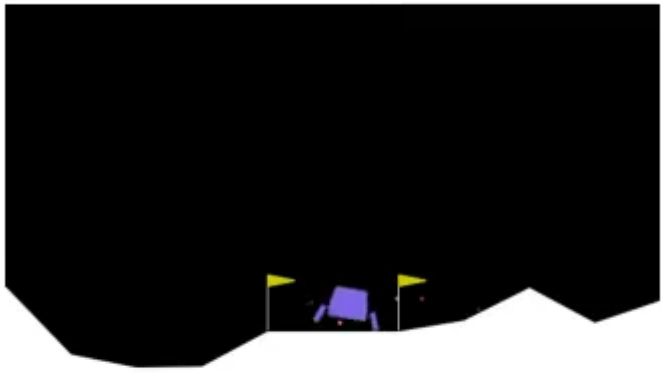
```
Episode 1100    Average Score: 111.85
Episode 1199    Average Score: 176.37
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1200    Average Score: 176.25
```



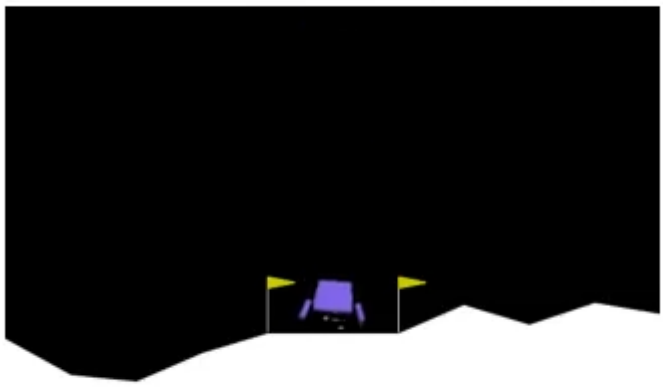
```
Episode 1200    Average Score: 176.25
Episode 1299    Average Score: 209.05
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1300    Average Score: 209.59
```



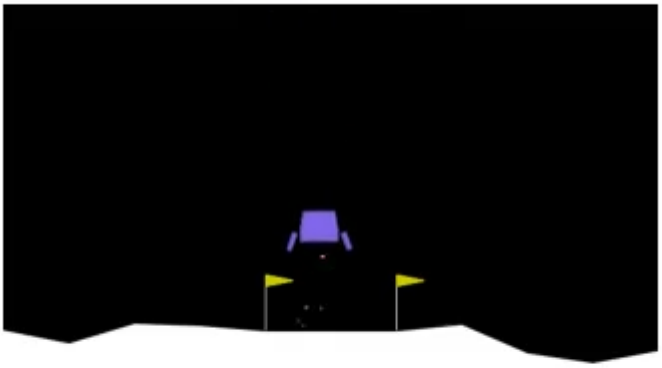
```
Episode 1300    Average Score: 209.59
Episode 1399    Average Score: 239.15
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1400    Average Score: 239.37
```



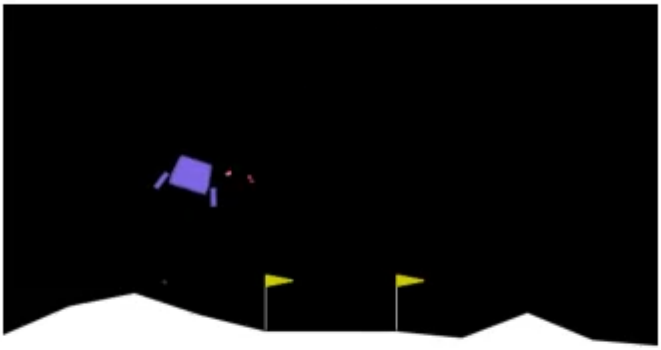
```
Episode 1400    Average Score: 239.37
Episode 1499    Average Score: 229.73
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1500    Average Score: 229.78
```



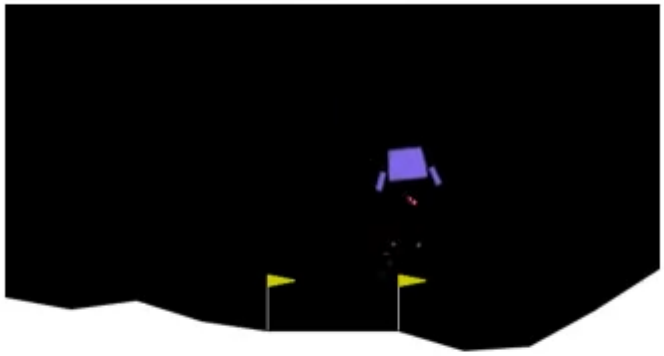
```
Episode 1500    Average Score: 229.78
Episode 1599    Average Score: 232.79
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1600    Average Score: 232.79
```



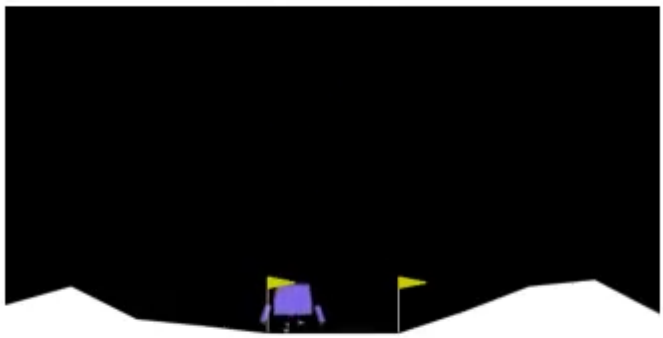
```
Episode 1600    Average Score: 232.79
Episode 1699    Average Score: 202.35
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1700    Average Score: 201.79
```



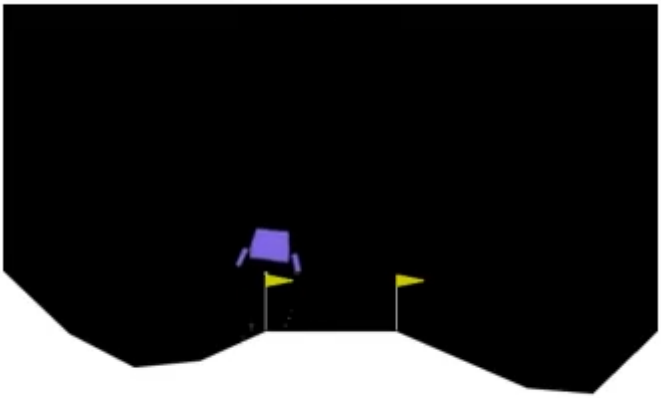
```
Episode 1700    Average Score: 201.79
Episode 1799    Average Score: 236.40
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1800    Average Score: 237.12
```



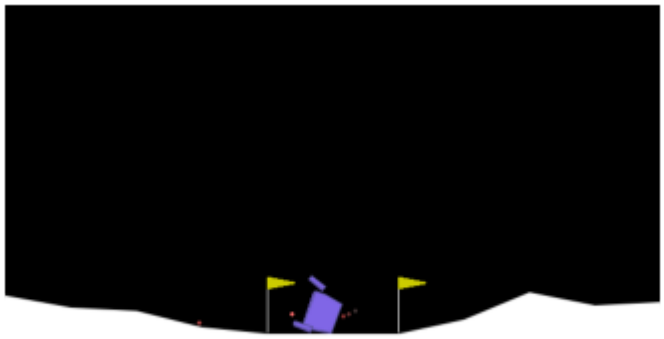
```
Episode 1800    Average Score: 237.12
Episode 1899    Average Score: 226.48
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1900    Average Score: 226.15
```

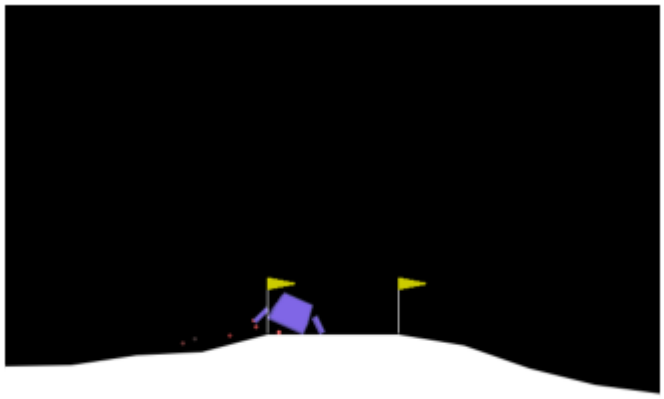


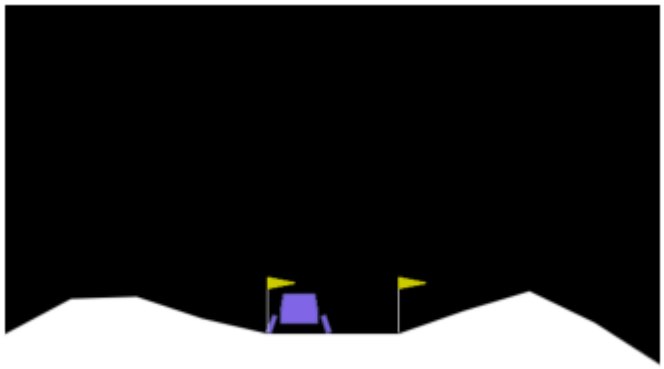
```
Episode 1900    Average Score: 226.15
Episode 1999    Average Score: 226.14
/opt/conda/lib/python3.8/site-packages/gym/core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2000    Average Score: 226.05
```

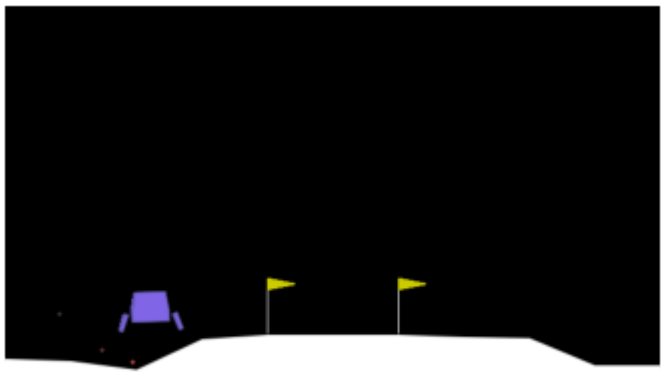


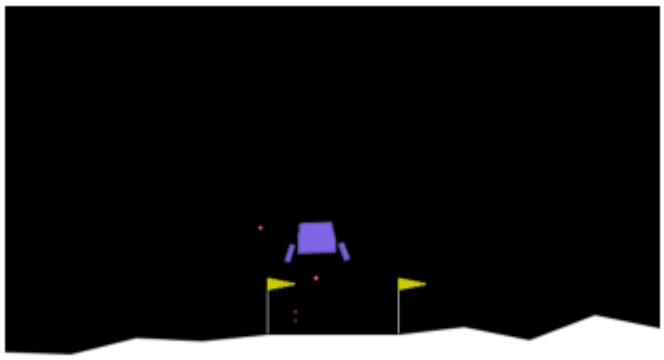
Episode 2000 Average Score: 226.05

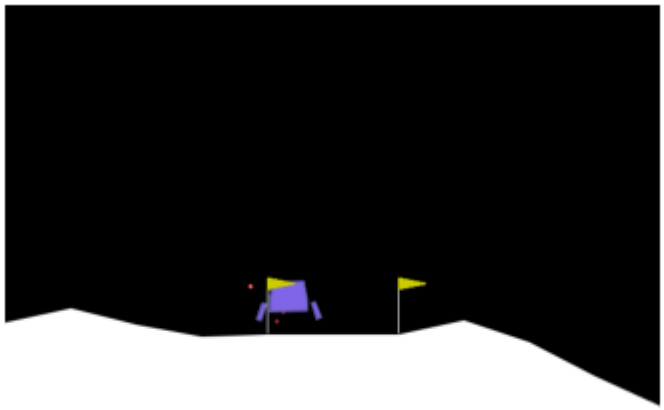


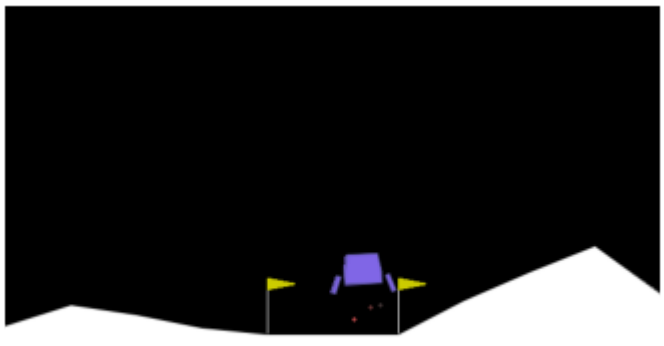


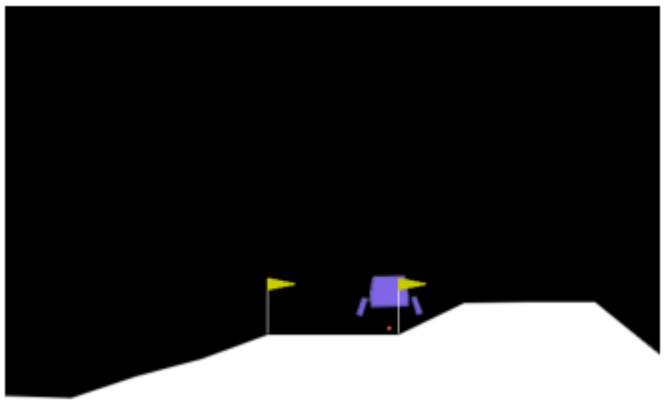


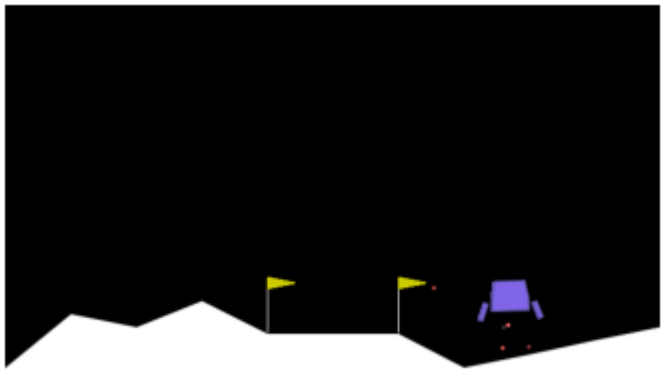


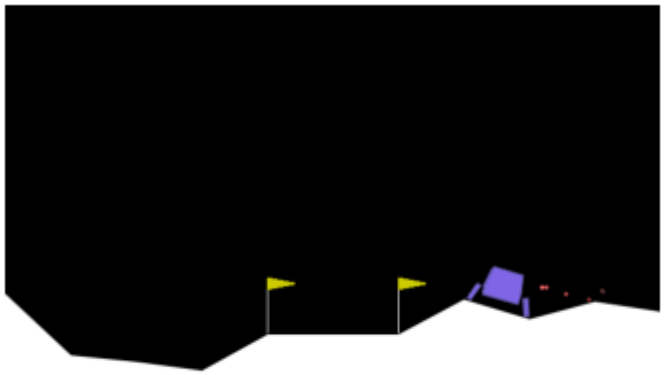


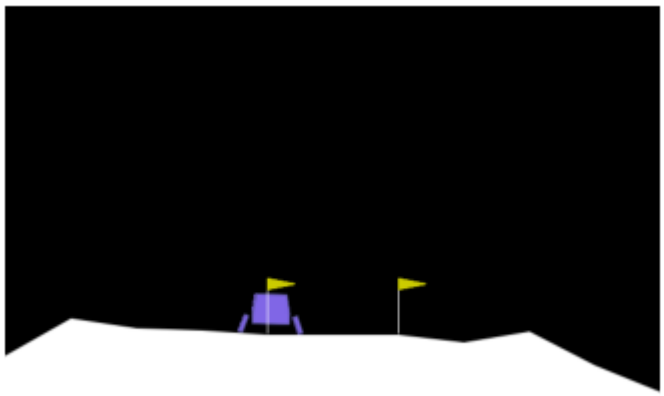


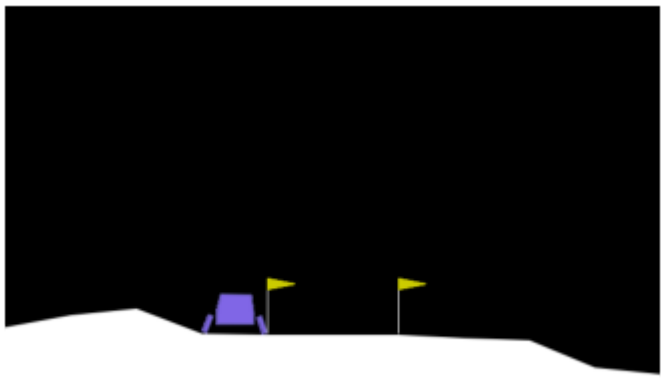


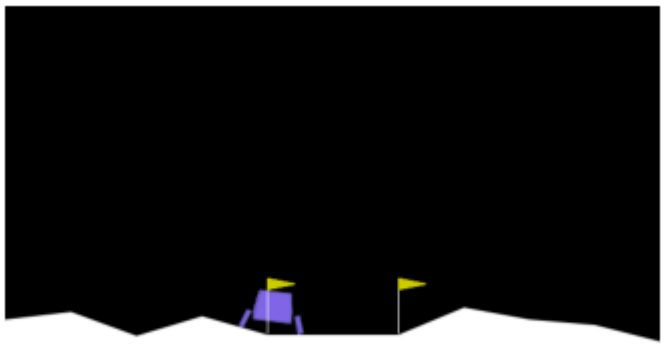


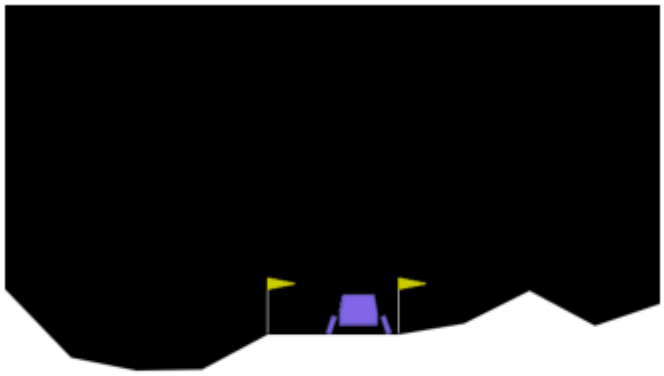


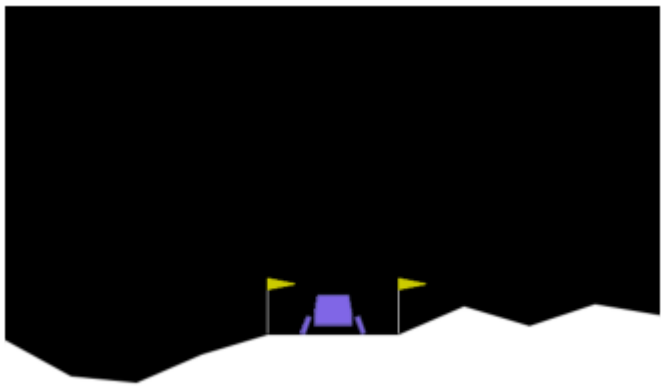


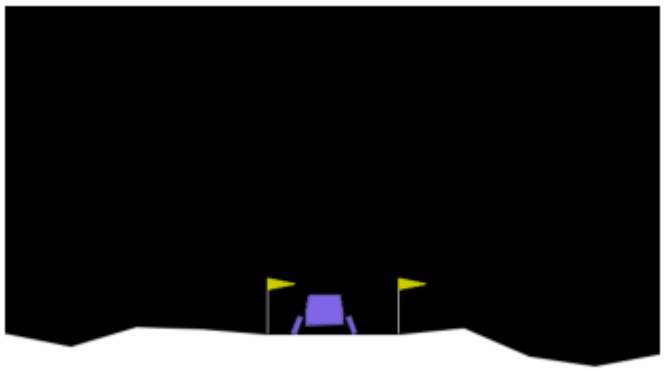


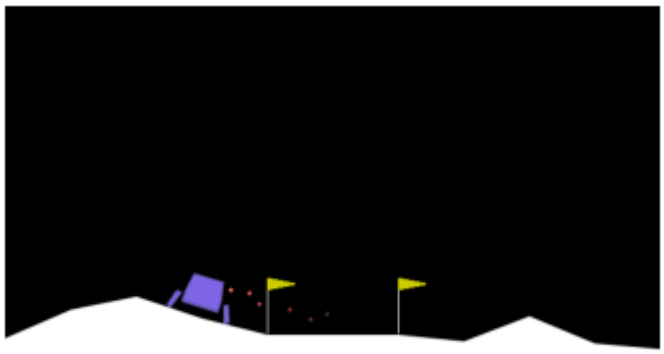


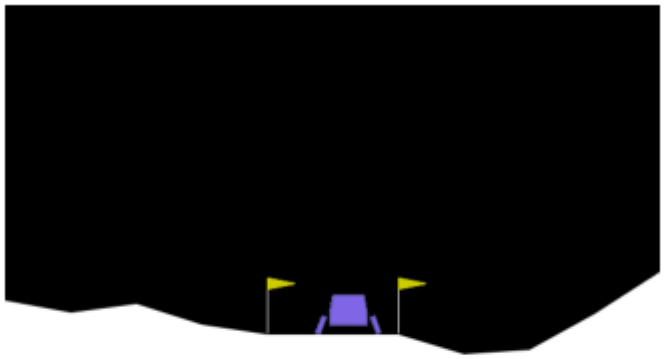


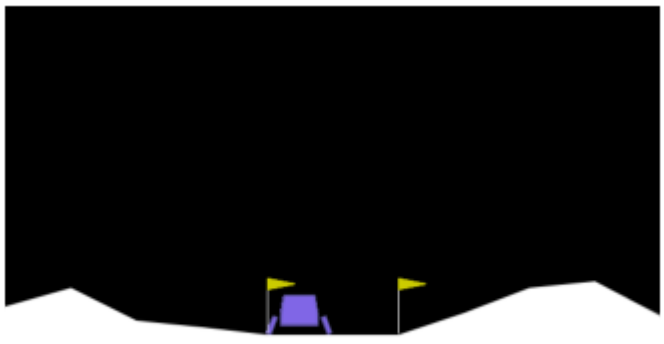








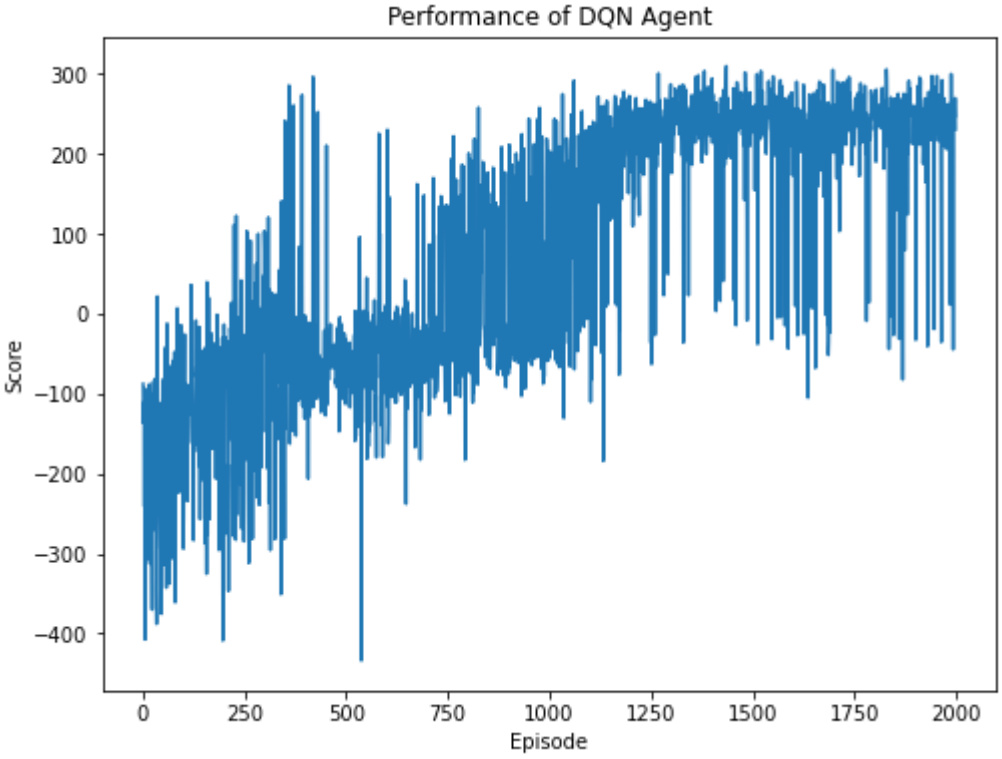




Model Improvement of Improved QNetwork

In [ ]:
```
fig = plt.figure(figsize=(8,6))
ax = fig.add_subplot(111)
plt.plot(np.arange(len(scores)), scores)
plt.ylabel('Score')
plt.xlabel('Episode')
plt.title('Performance of DQN Agent 2')
plt.show()
```

```
/opt/conda/lib/python3.8/site-packages/ipykernel/ipkernel.py:287: DeprecationWarning: `should
_run_async` will not call `transform_cell` automatically in the future. Please pass the resul
t to `transformed_cell` argument and any exception that happen during thetransform in `prepro
cessing_exc_tuple` in IPython 7.17 and above.
  and should_run_async(code)
```



Observations

The added hidden layer turned out to be extremely effective.
Our score from the first model was struggling to hit a score of 200 and above.
It now hits a score of 200 as early as episode 400 and continues to rise and hover around 200 after episode 1000.

DQN Hypertuning

Increased Epsilon by 0.5

In [ ]:
```
class QNetwork_3(nn.Module):

    def __init__(self, state_size, action_size, seed):
        """Initialize parameters and build model.
        Params
        ======
            state_size (int): Dimension of each state
            action_size (int): Dimension of each action
            seed (int): Random seed
        """
        super(QNetwork_3, self).__init__()
        self.seed = torch.manual_seed(seed)
        self.fc1 = nn.Linear(state_size, 128)
        self.fc2 = nn.Linear(128, 128)
        self.fc3 = nn.Linear(128, 128)
        self.fc4 = nn.Linear(128, action_size)

    def forward(self, state):
        """Build a network that maps state -> action values."""
        x = self.fc1(state)
        x = F.relu(x)
        x = self.fc2(x)
        x = F.relu(x)
        x = self.fc3(x)
        x = F.relu(x)
        return self.fc4(x)
```

In [ ]:
```
class DQNAgent_3():

    def __init__(self, state_size, action_size, seed):

        self.state_size = state_size
        self.action_size = action_size
        self.seed = random.seed(seed)

        self.qnetwork_local = QNetwork_3(state_size, action_size, seed).to(device)
        self.qnetwork_target = QNetwork_3(state_size, action_size, seed).to(device)
        self.optimizer = optim.Adam(self.qnetwork_local.parameters(), lr=LR)

        self.memory = ReplayBuffer(action_size, BUFFER_SIZE, BATCH_SIZE, seed)
        self.t_step = 0

    def step(self, state, action, reward, next_state, done):
        # Save experience in replay memory
        self.memory.new_experience(state, action, reward, next_state, done)

        # Learn every UPDATE_EVERY time steps.
        self.t_step = (self.t_step + 1) % UPDATE_EVERY
        if self.t_step == 0:
            # If enough samples are available in memory, get random subset and learn
            if len(self.memory) > BATCH_SIZE:
                experiences = self.memory.sample()
                self.learn(experiences, GAMMA)

    def act(self, state, eps=0.):

        state = torch.from_numpy(state).float().unsqueeze(0).to(device)
        self.qnetwork_local.eval()
        with torch.no_grad():
            action_values = self.qnetwork_local(state)
        self.qnetwork_local.train()
```

```
        # Epsilon-greedy action selection
        if random.random() > eps:
            return np.argmax(action_values.cpu().data.numpy())
        else:
            return random.choice(np.arange(self.action_size))

    def learn(self, experiences, gamma):

        # Obtain random minibatch of tuples from D
        states, actions, rewards, next_states, dones = experiences

        ## Compute and minimize the loss
        q_targets_next = self.qnetwork_target(next_states).detach().max(1)[0].unsqueeze(1)
        q_targets = rewards + gamma * q_targets_next * (1 - dones)
        q_expected = self.qnetwork_local(states).gather(1, actions)

        ### Loss calculation (we used Mean squared error)
        loss = F.mse_loss(q_expected, q_targets)
        self.optimizer.zero_grad()
        loss.backward()
        self.optimizer.step()

        # ------------------- update target network ------------------- #
        self.soft_update(self.qnetwork_local, self.qnetwork_target, TAU)

    def soft_update(self, local_model, target_model, tau):

        for target_param, local_param in zip(target_model.parameters(), local_model.parameter
            target_param.data.copy_(tau*local_param.data + (1.0-tau)*target_param.data)
```

In [ ]:

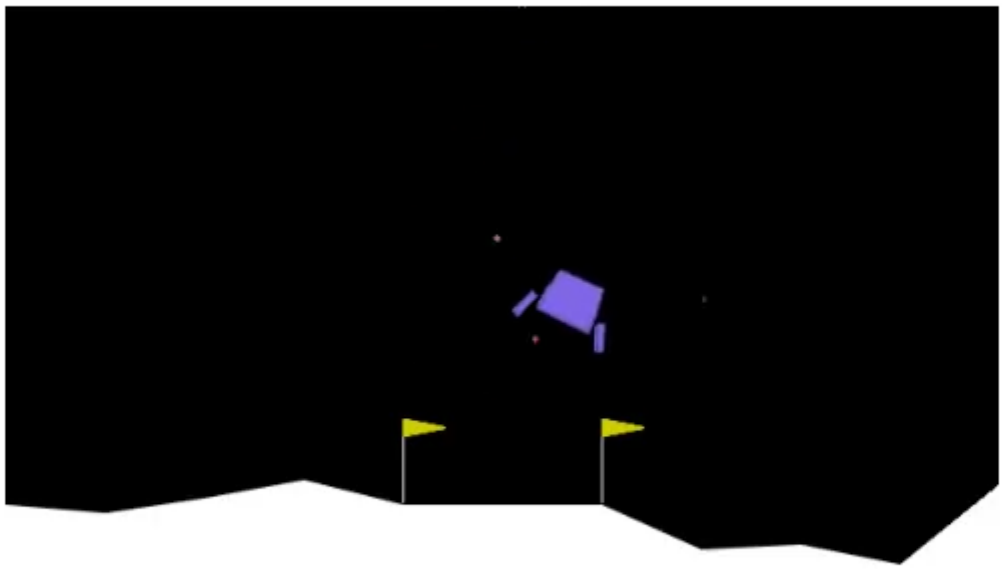
```
def train_2(n_episodes=5000, max_t=1000, eps_start=1.5, eps_end=0.01, eps_decay=0.995):

    master_frames = []
    scores = []
    scores_window = deque(maxlen=100)
    eps = eps_start
    for i_episode in range(1, n_episodes+1):
        state = env.reset()
        score = 0
        if i_episode % 100 == 0:
            frames = []
        for t in range(max_t):
            action = agent.act(state, eps)
            next_state, reward, done, _ = env.step(action)
            agent.step(state, action, reward, next_state, done)
            if i_episode % 100 == 0:
                screen = env.render(mode='rgb_array')
                frames.append(screen)
            state = next_state
            score += reward
            if done:
                break
        scores_window.append(score)       # save most recent score
        scores.append(score)              # save most recent score
        eps = max(eps_end, eps_decay*eps) # decrease epsilon
        print('\rEpisode {}\tAverage Score: {:.2f}'.format(i_episode, np.mean(scores_window))
        if i_episode % 100 == 0:
            create_animation(frames)
            master_frames.append(frames)
            print('\rEpisode {}\tAverage Score: {:.2f}'.format(i_episode, np.mean(scores_wind
        torch.save(agent.qnetwork_local.state_dict(), 'checkpoint.pth')
    return scores, master_frames
```

In [ ]:

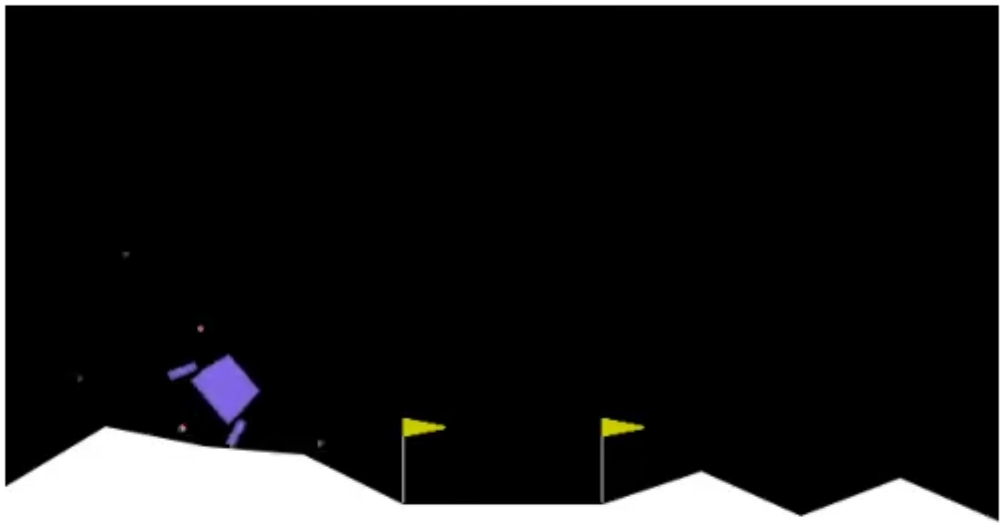
```
agent = DQNAgent_3(state_size=8, action_size=4, seed=0)
scores, master_frames = train_2()
```

```
Episode 99      Average Score: -168.57
```

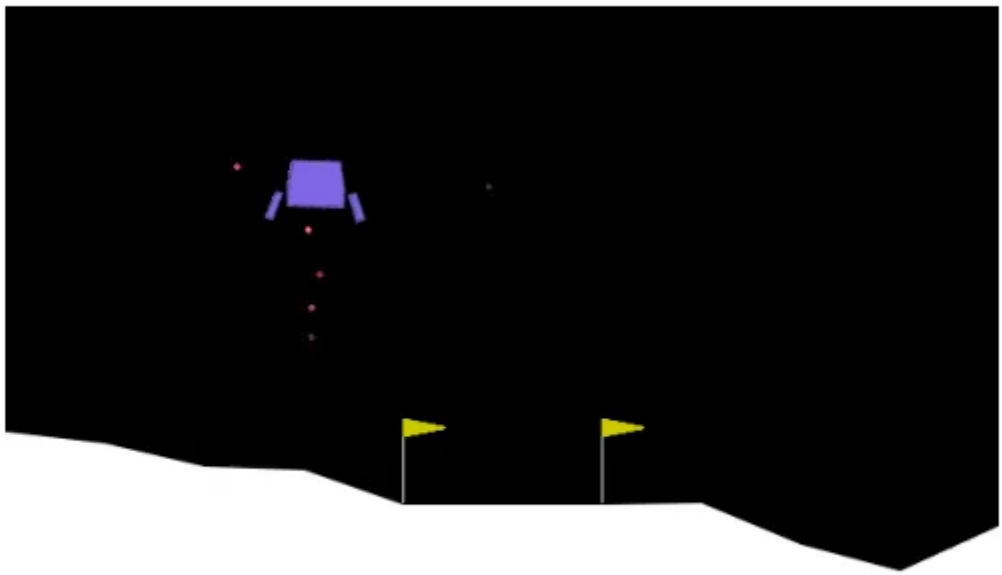
```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 100     Average Score: -168.58
```



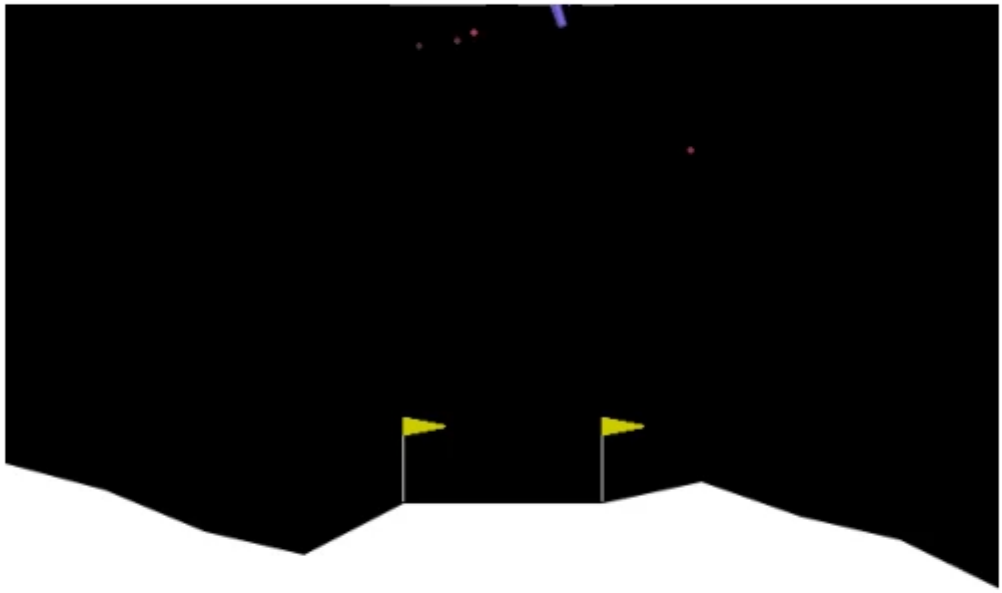
```
Episode 100     Average Score: -168.58
Episode 199     Average Score: -148.68
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 200     Average Score: -147.73
```



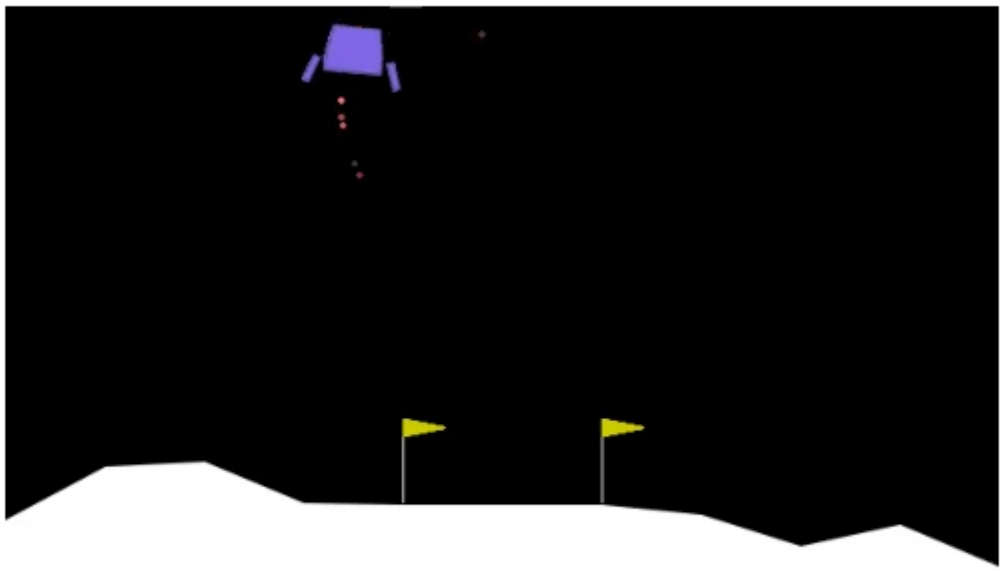
```
Episode 200     Average Score: -147.73
Episode 299     Average Score: -131.52
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 300     Average Score: -131.36
```



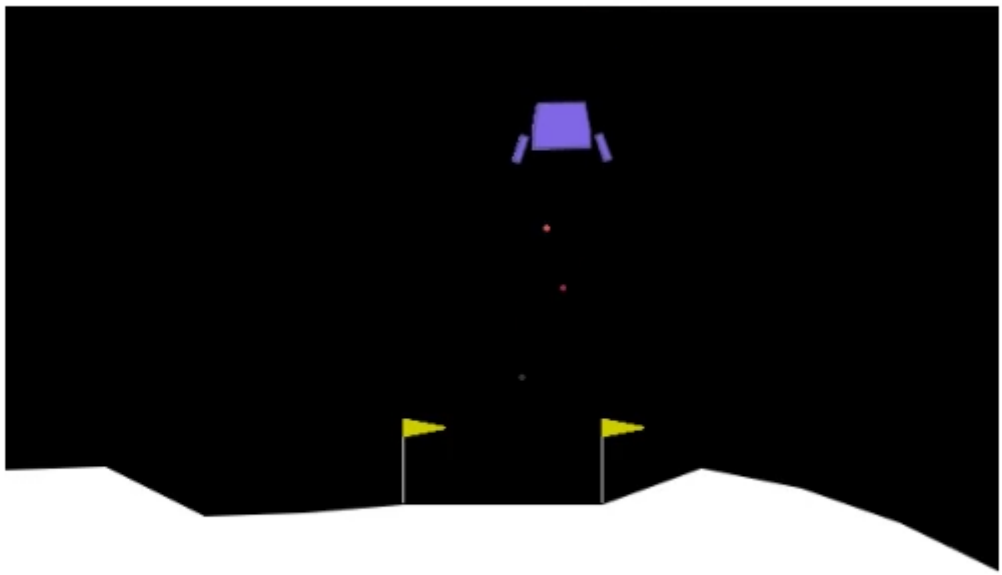
```
Episode 300     Average Score: -131.36
Episode 399     Average Score: -106.96
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 400     Average Score: -106.53
```



```
Episode 400     Average Score: -106.53
Episode 499     Average Score: -70.368
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 500     Average Score: -70.69
```



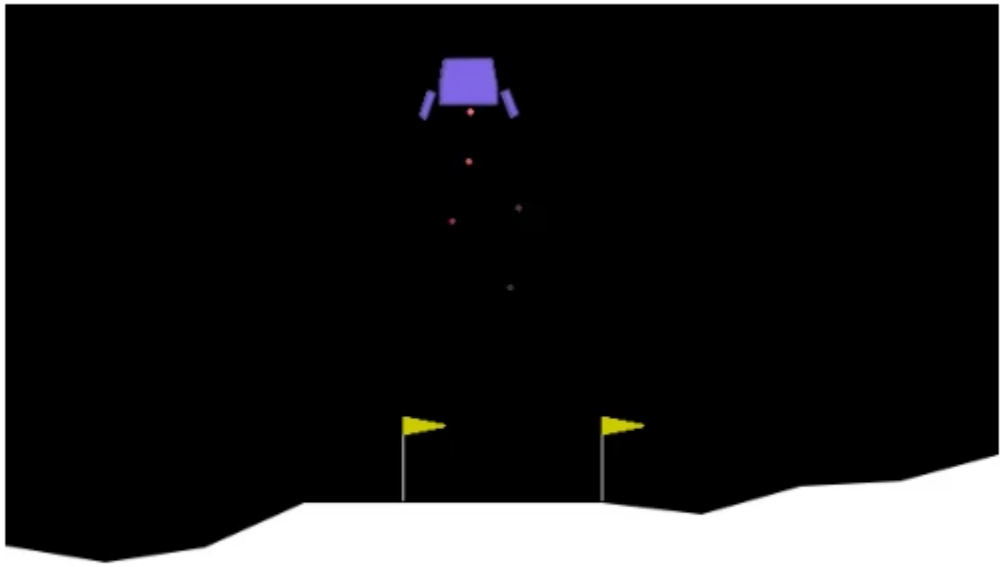
```
Episode 500     Average Score: -70.69
Episode 599     Average Score: -50.86
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 600     Average Score: -50.46
```



```
Episode 600     Average Score: -50.46
Episode 699     Average Score: -34.56
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

```
Episode 700     Average Score: -34.85
```

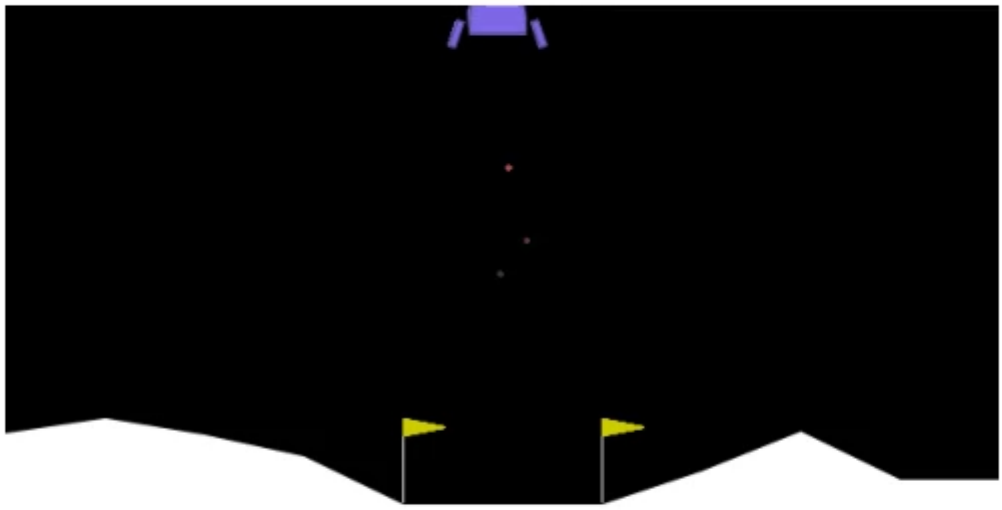



```
Episode 700     Average Score: -34.85
```

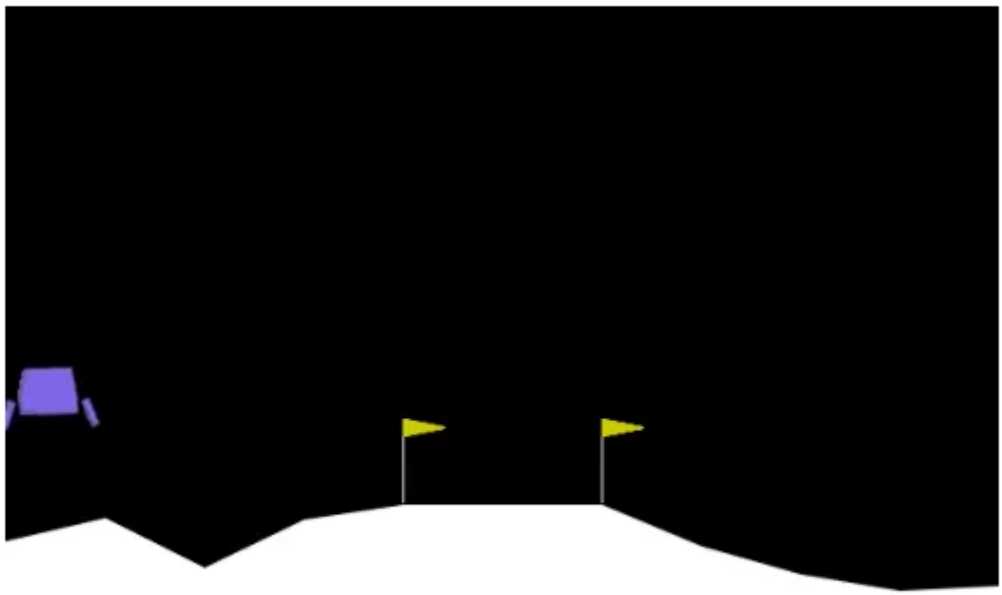

```
Episode 799     Average Score: -33.05
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

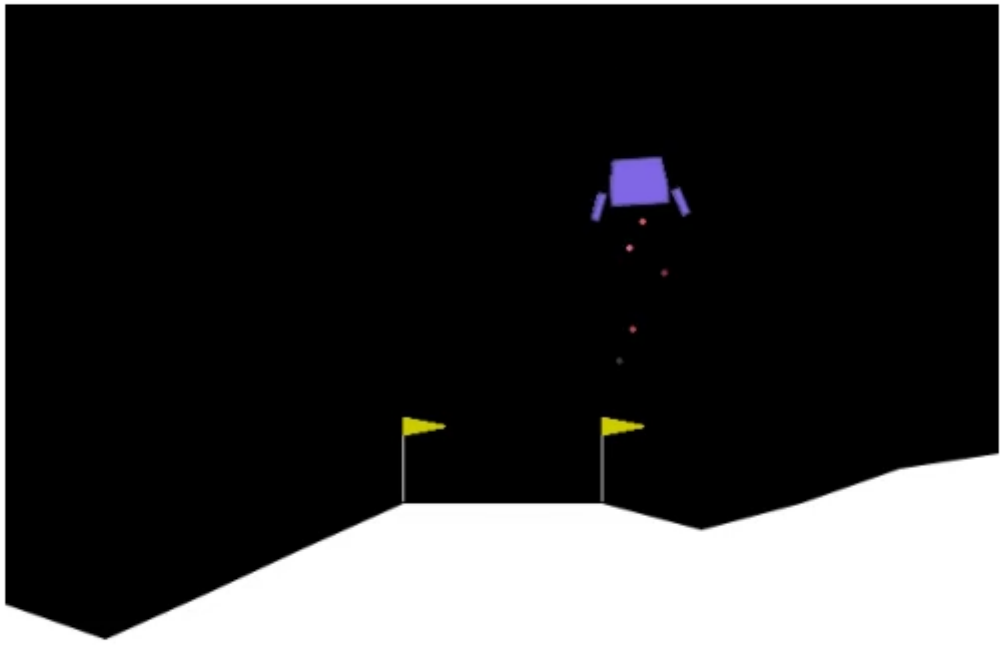
```
Episode 800     Average Score: -32.97
```



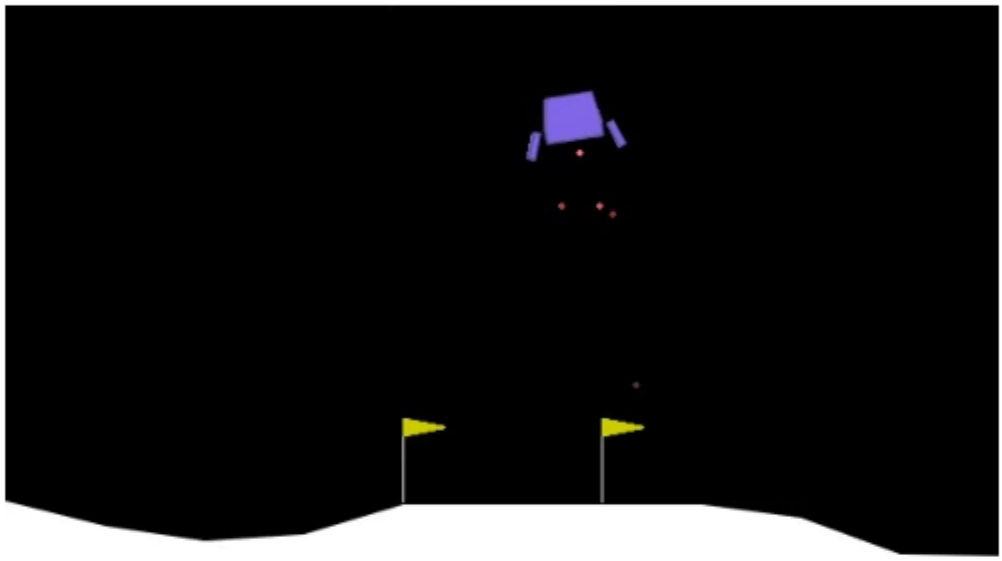
```
Episode 800     Average Score: -32.97
Episode 899     Average Score: -12.06
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 900     Average Score: -13.15
```



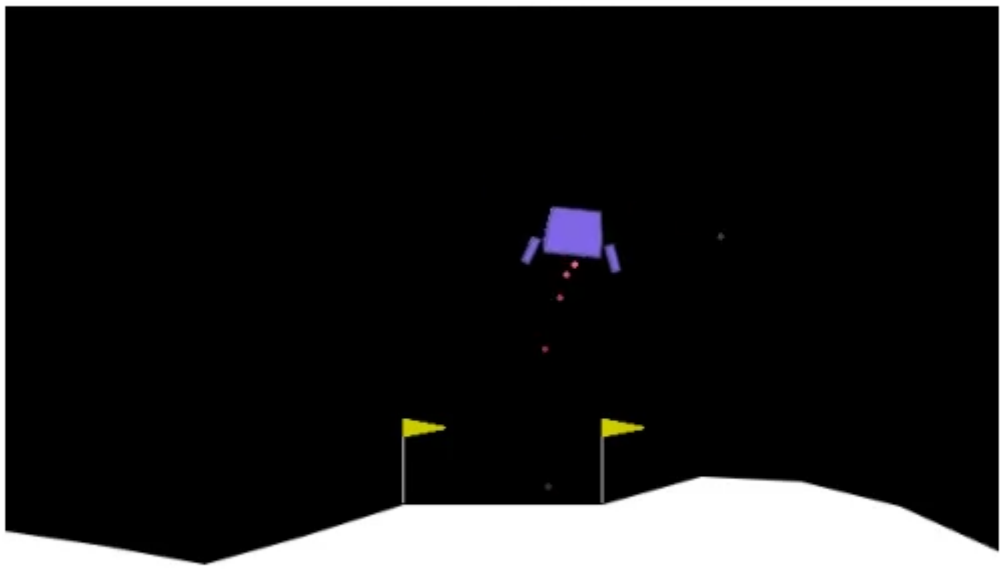
```
Episode 900     Average Score: -13.15
Episode 999     Average Score: 14.973
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1000    Average Score: 16.31
```



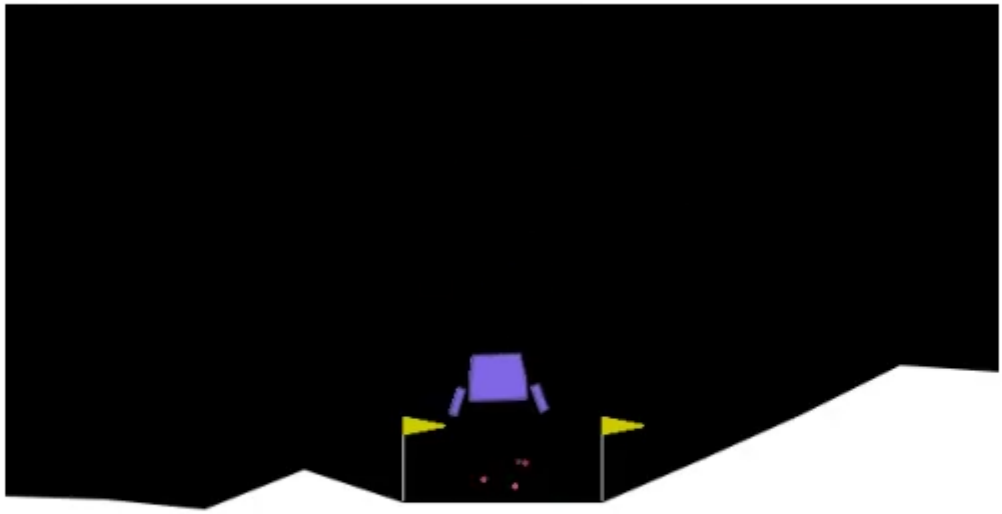
```
Episode 1000    Average Score: 16.31
Episode 1099    Average Score: 32.38
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1100    Average Score: 33.99
```



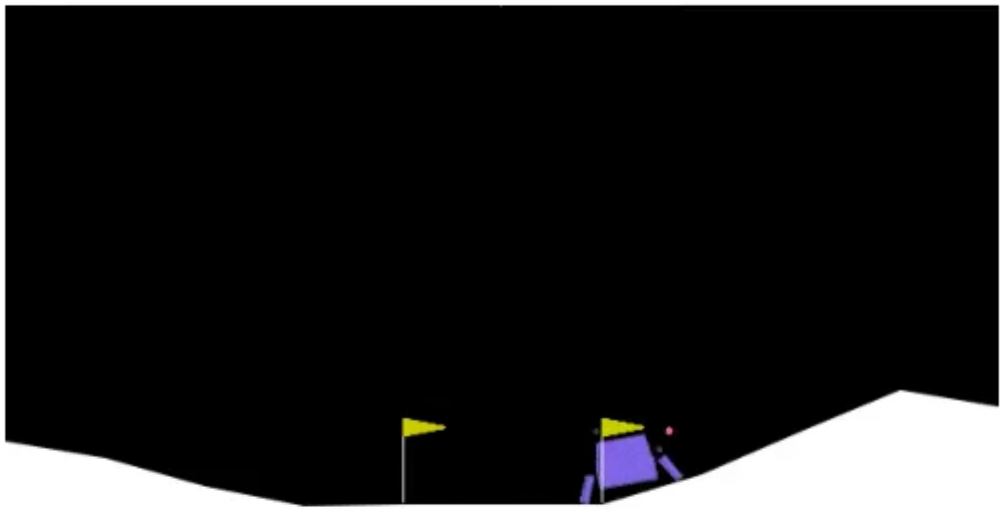
```
Episode 1100    Average Score: 33.99
Episode 1199    Average Score: 128.48
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1200    Average Score: 128.70
```



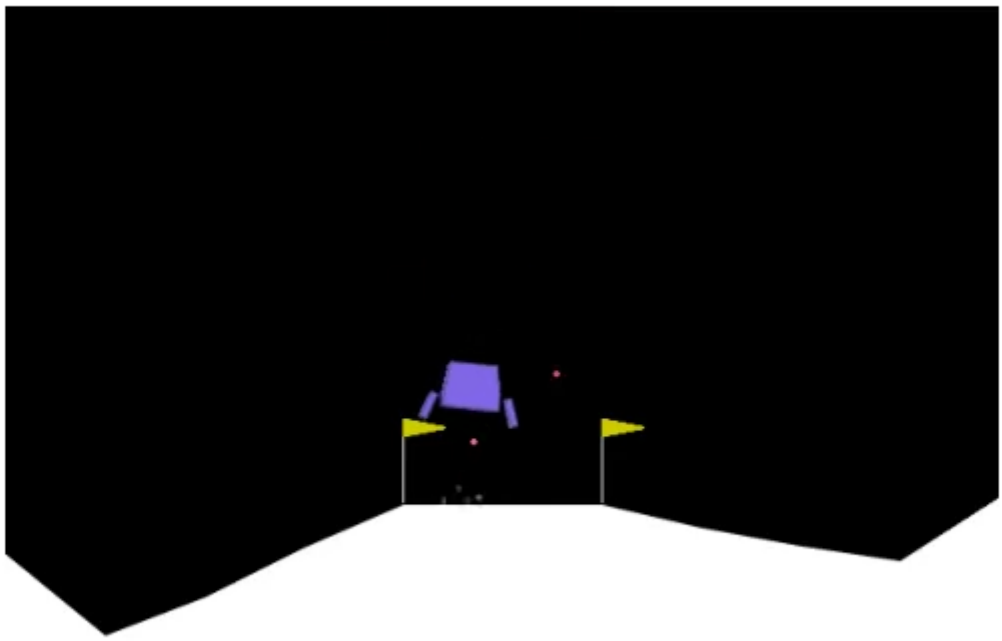
```
Episode 1200    Average Score: 128.70
Episode 1299    Average Score: 165.10
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1300    Average Score: 163.50
```



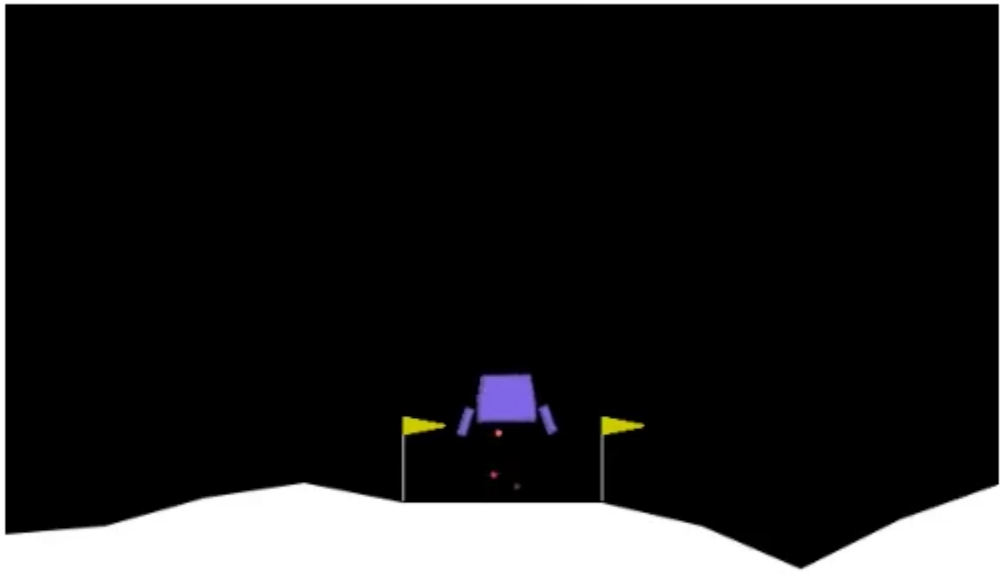
```
Episode 1300    Average Score: 163.50
Episode 1399    Average Score: 194.52
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1400    Average Score: 196.86
```



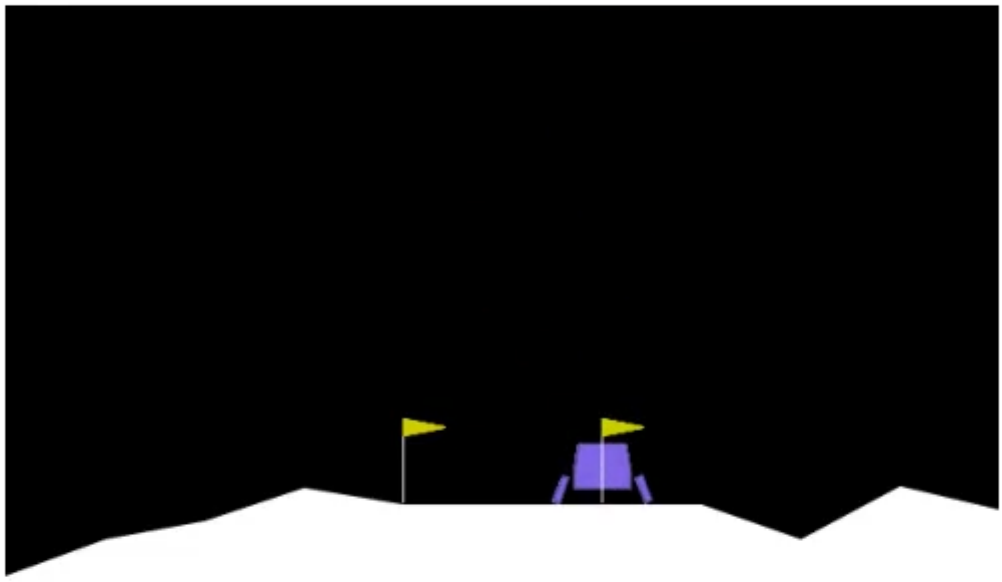
```
Episode 1400    Average Score: 196.86
Episode 1499    Average Score: 222.65
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1500    Average Score: 222.25
```



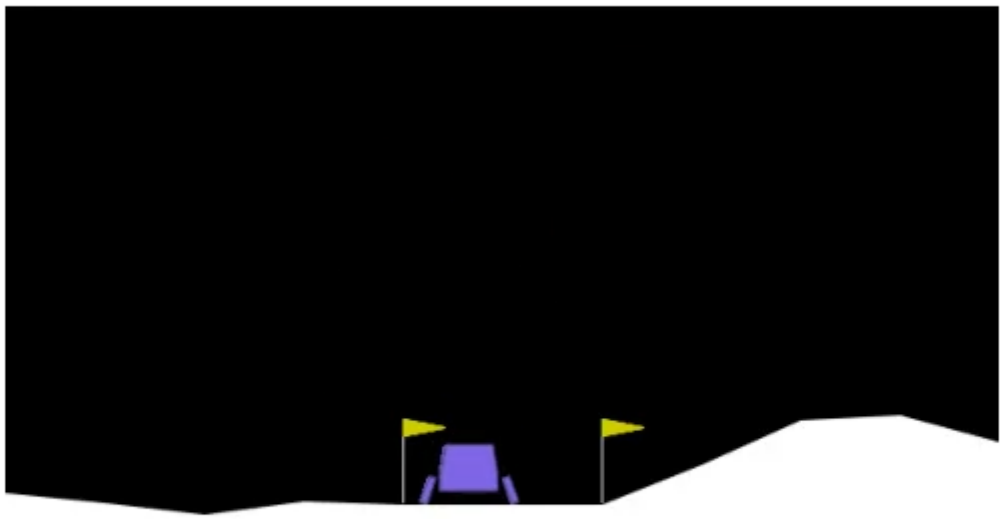
```
Episode 1500    Average Score: 222.25
Episode 1599    Average Score: 229.48
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1600    Average Score: 229.82
```



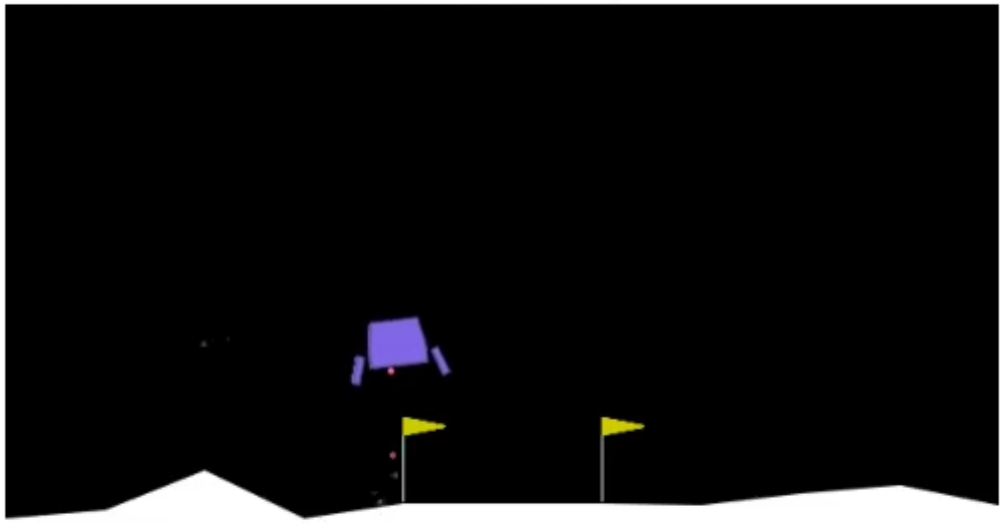
```
Episode 1600    Average Score: 229.82
Episode 1699    Average Score: 243.29
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1700    Average Score: 243.59
```



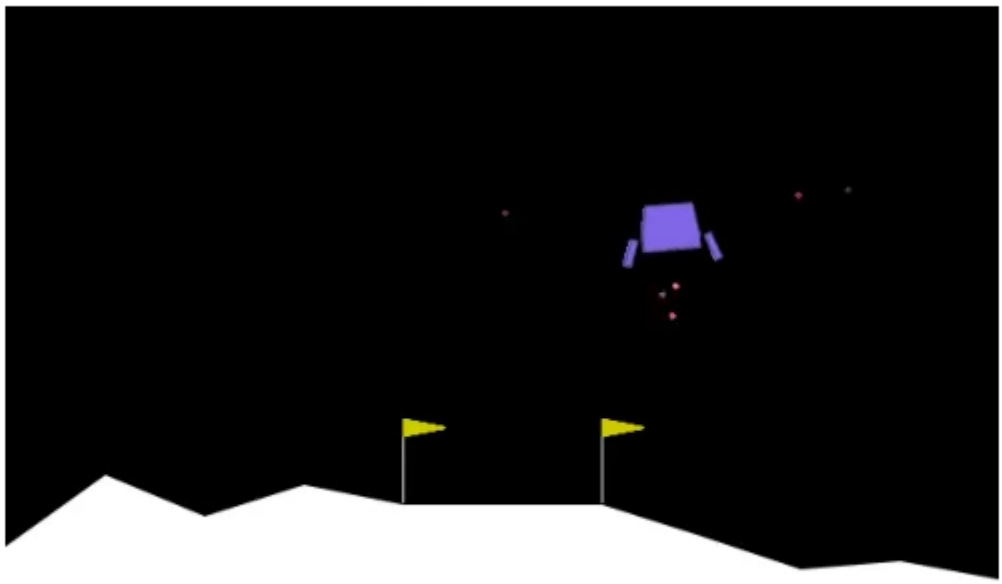
```
Episode 1700    Average Score: 243.59
Episode 1799    Average Score: 251.36
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1800    Average Score: 251.11
```



```
Episode 1800    Average Score: 251.11
Episode 1899    Average Score: 246.95
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 1900    Average Score: 247.23
```



```
Episode 1900    Average Score: 247.23
Episode 1999    Average Score: 240.50
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2000    Average Score: 240.46
```

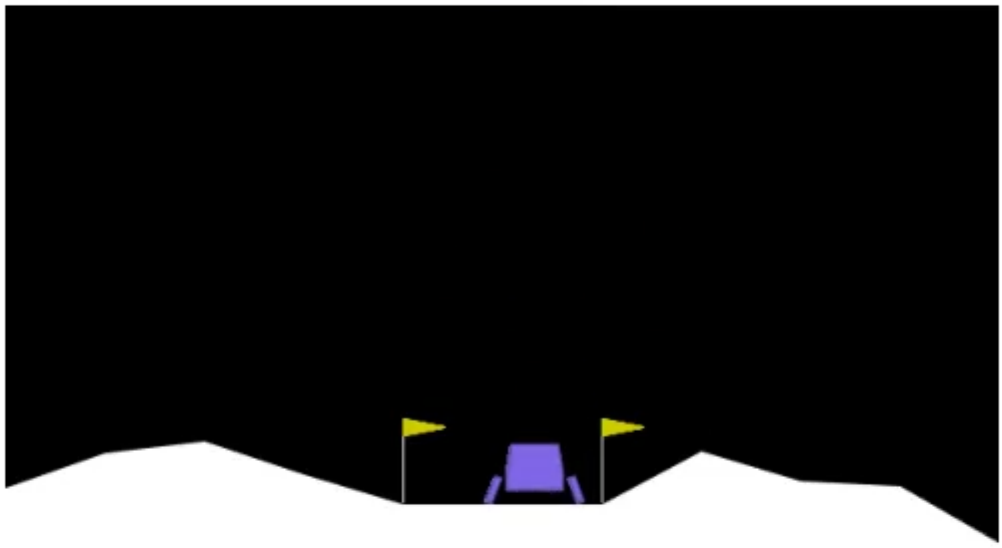


```
Episode 2000    Average Score: 240.46
Episode 2099    Average Score: 242.84
```

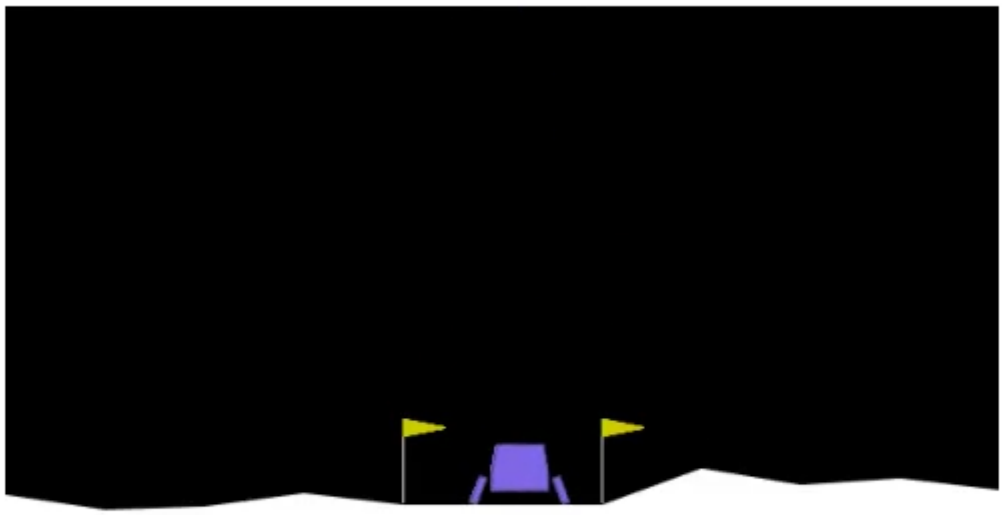
```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

```
Episode 2100    Average Score: 242.83
```

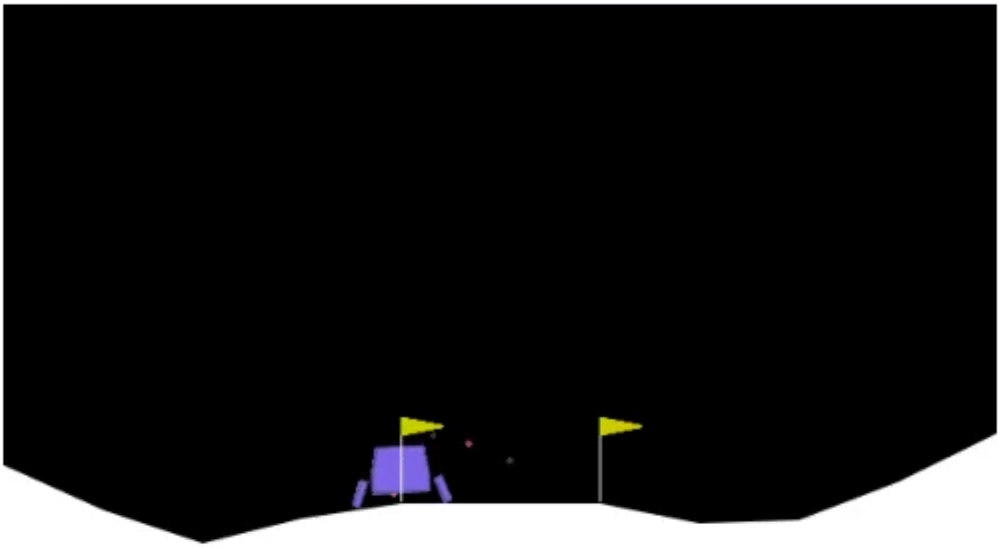
```
C:\Users\Soh Hong Yu\AppData\Local\Temp\ipykernel_30964\3541388167.py:3: RuntimeWarning: More than 20 figures have been opened. Figures created through the pyplot interface (`matplotlib.pyplot.figure`) are retained until explicitly closed and may consume too much memory. (To control this warning, see the rcParam `figure.max_open_warning`). Consider using `matplotlib.pyplot.close()`.
  fig = plt.figure()
```



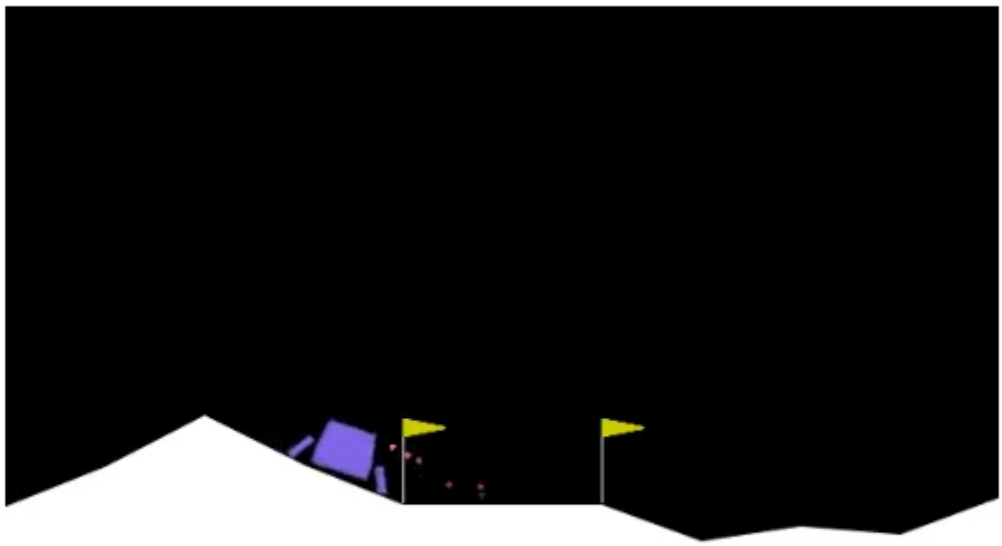
```
Episode 2100    Average Score: 242.83
Episode 2199    Average Score: 232.39
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2200    Average Score: 232.33
```



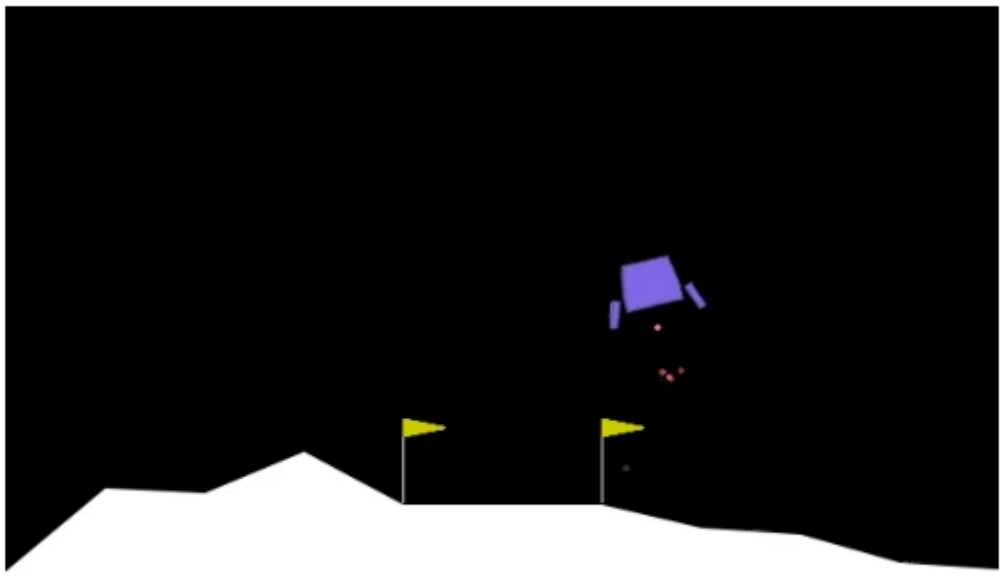
```
Episode 2200    Average Score: 232.33
Episode 2299    Average Score: 234.80
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2300    Average Score: 234.77
```



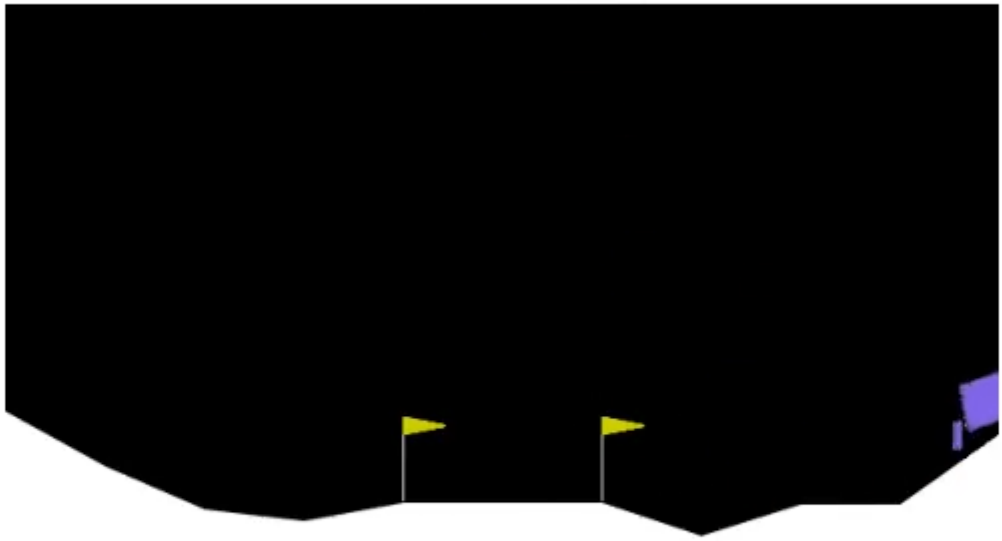
```
Episode 2300    Average Score: 234.77
Episode 2399    Average Score: 229.88
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2400    Average Score: 227.48
```



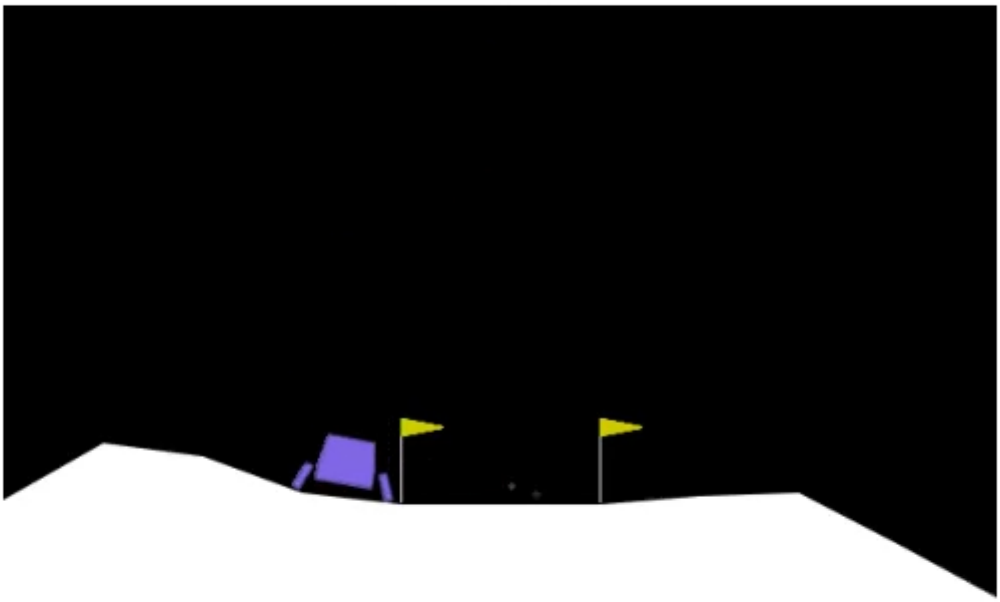
```
Episode 2400    Average Score: 227.48
Episode 2499    Average Score: 229.94
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2500    Average Score: 232.22
```



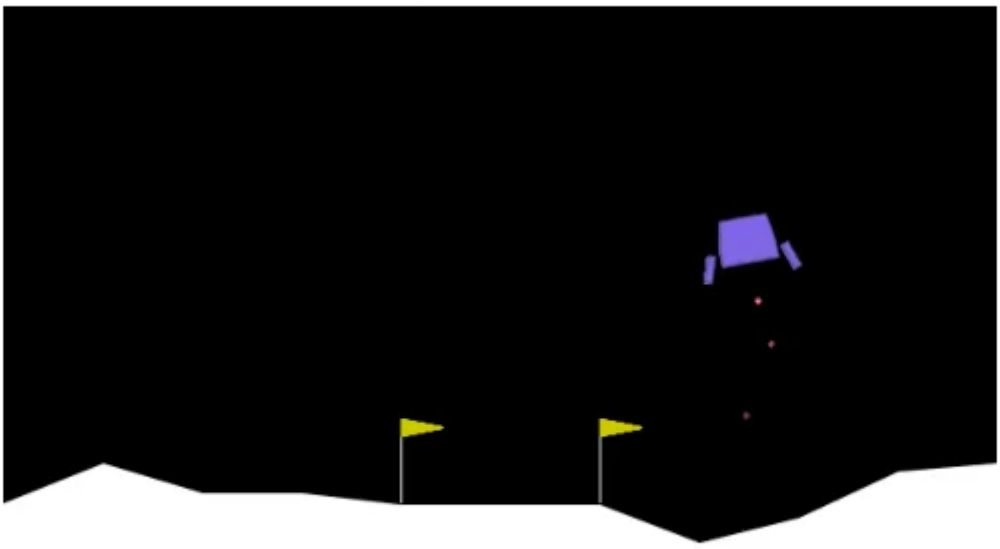
```
Episode 2500    Average Score: 232.22
Episode 2599    Average Score: 224.61
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2600    Average Score: 221.50
```



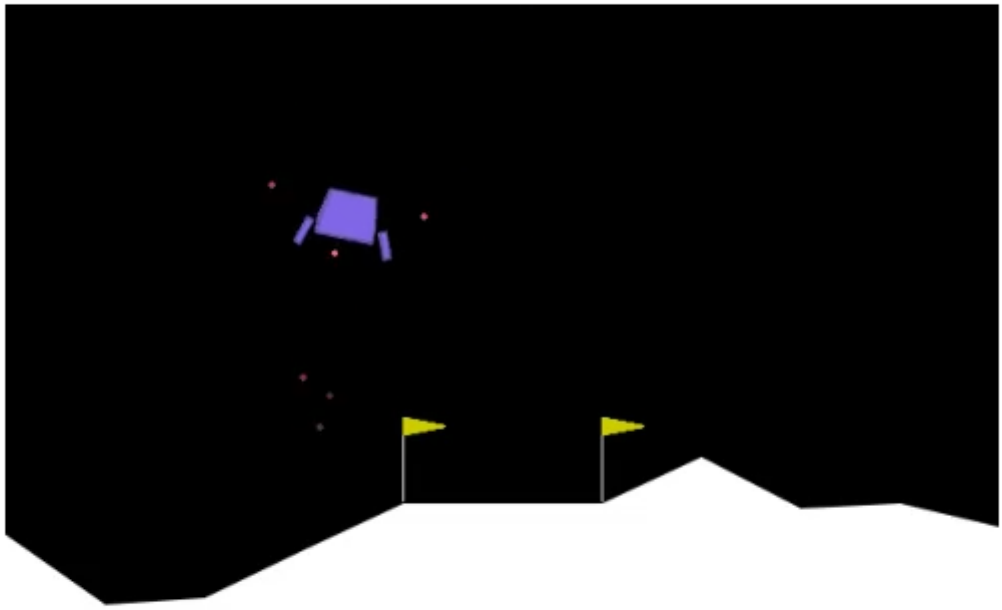
```
Episode 2600    Average Score: 221.50
Episode 2699    Average Score: 229.38
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2700    Average Score: 232.29
```



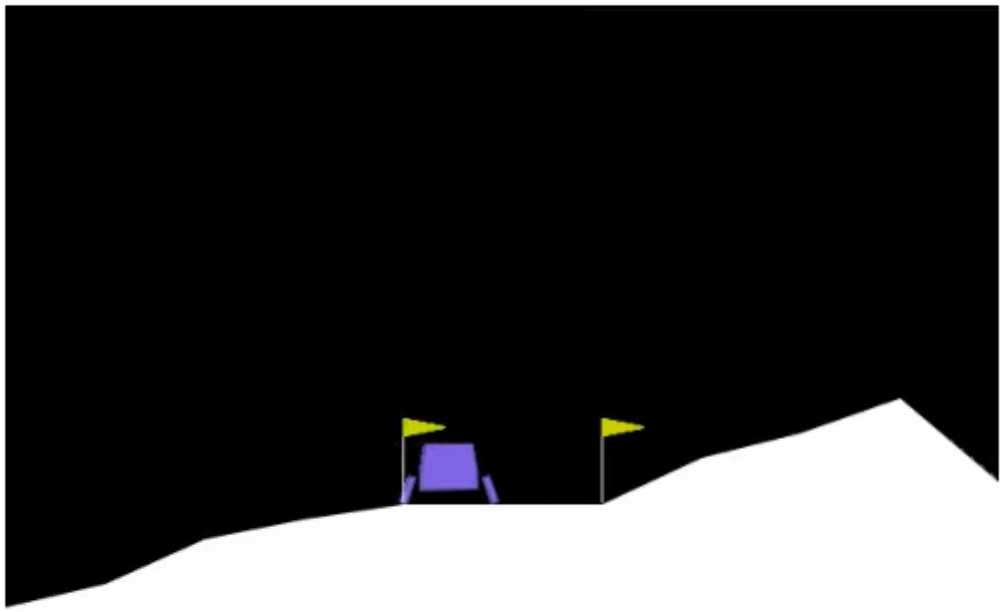
```
Episode 2700    Average Score: 232.29
Episode 2799    Average Score: 235.71
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2800    Average Score: 235.90
```



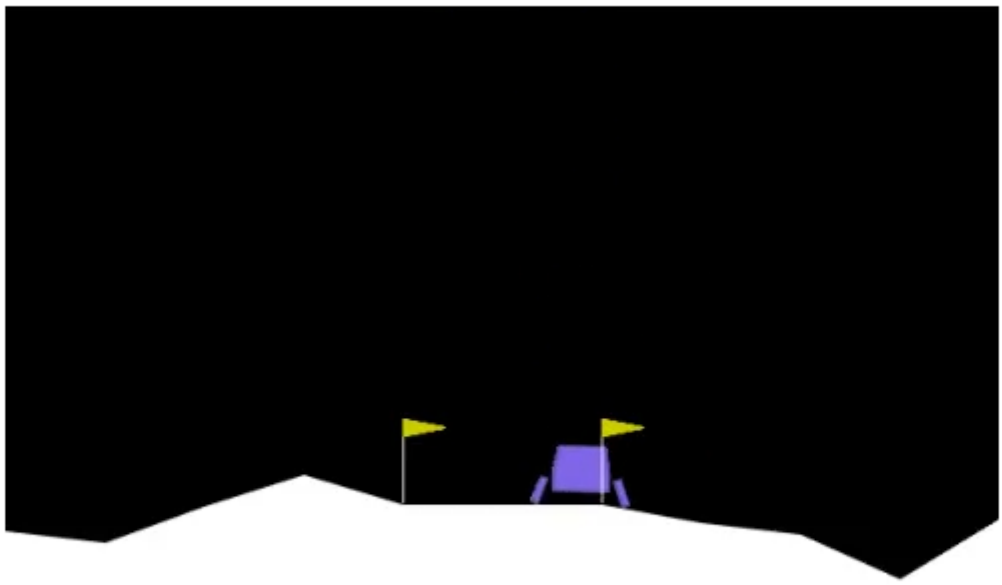
```
Episode 2800    Average Score: 235.90
Episode 2899    Average Score: 241.22
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 2900    Average Score: 239.99
```



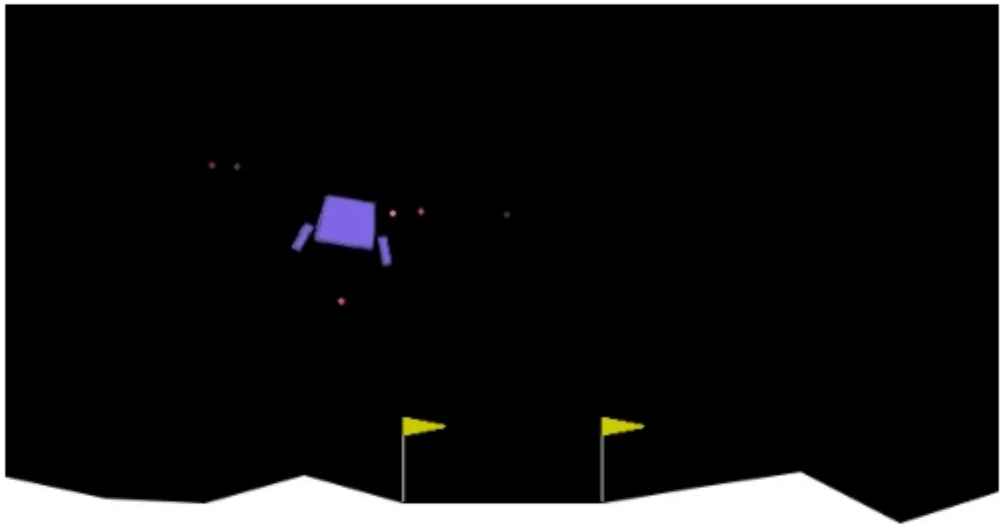
```
Episode 2900    Average Score: 239.99
Episode 2999    Average Score: 249.25
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3000    Average Score: 250.71
```



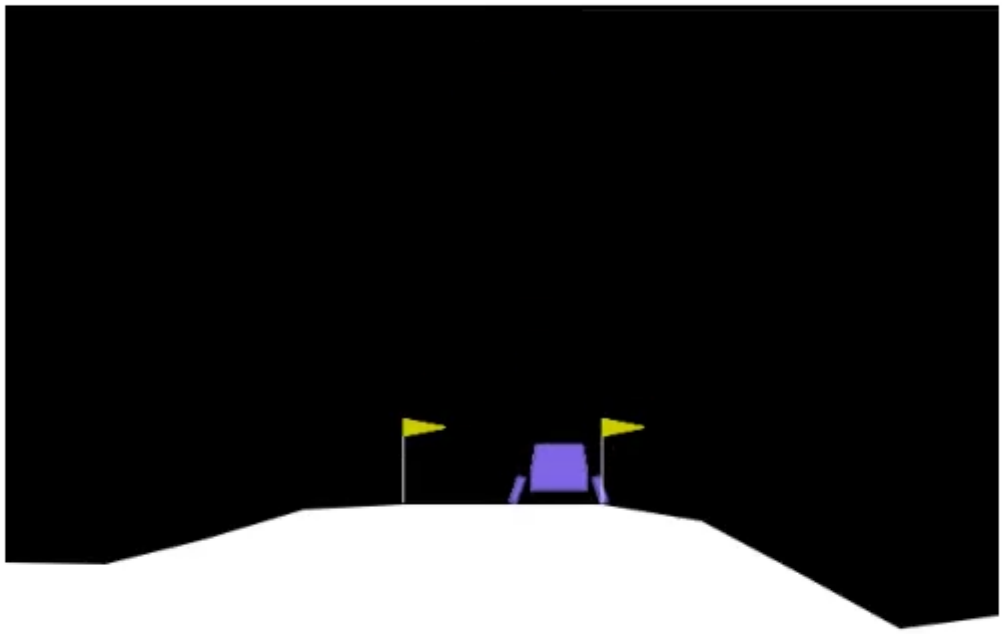
```
Episode 3000    Average Score: 250.71
Episode 3099    Average Score: 237.16
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3100    Average Score: 237.09
```



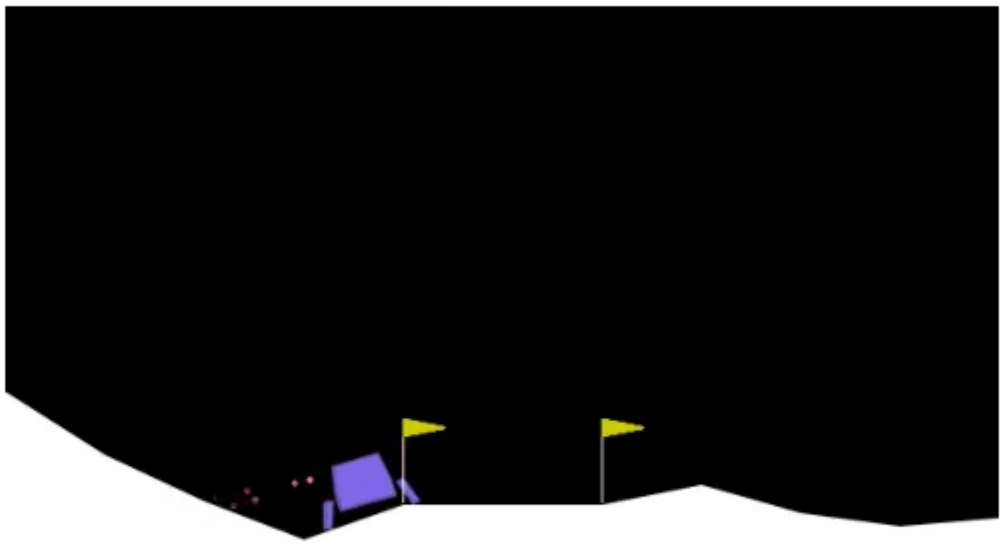
```
Episode 3100    Average Score: 237.09
Episode 3199    Average Score: 238.39
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3200    Average Score: 237.97
```



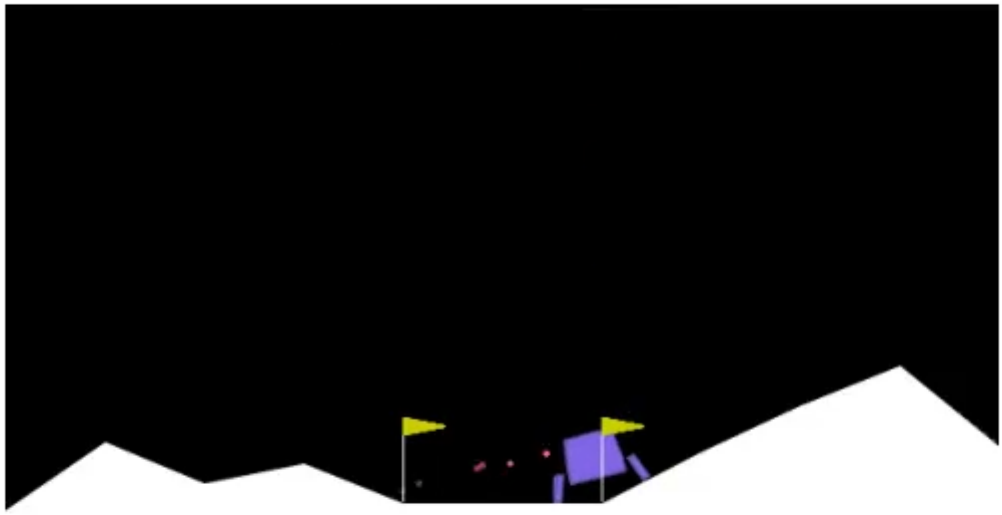
```
Episode 3200    Average Score: 237.97
Episode 3299    Average Score: 218.85
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3300    Average Score: 219.35
```



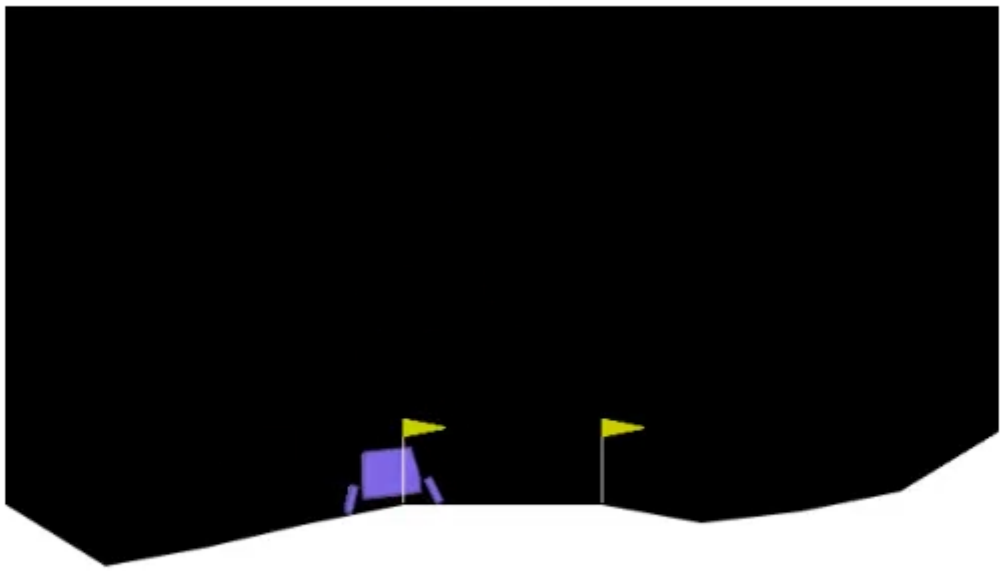
```
Episode 3300    Average Score: 219.35
Episode 3399    Average Score: 230.92
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3400    Average Score: 230.62
```



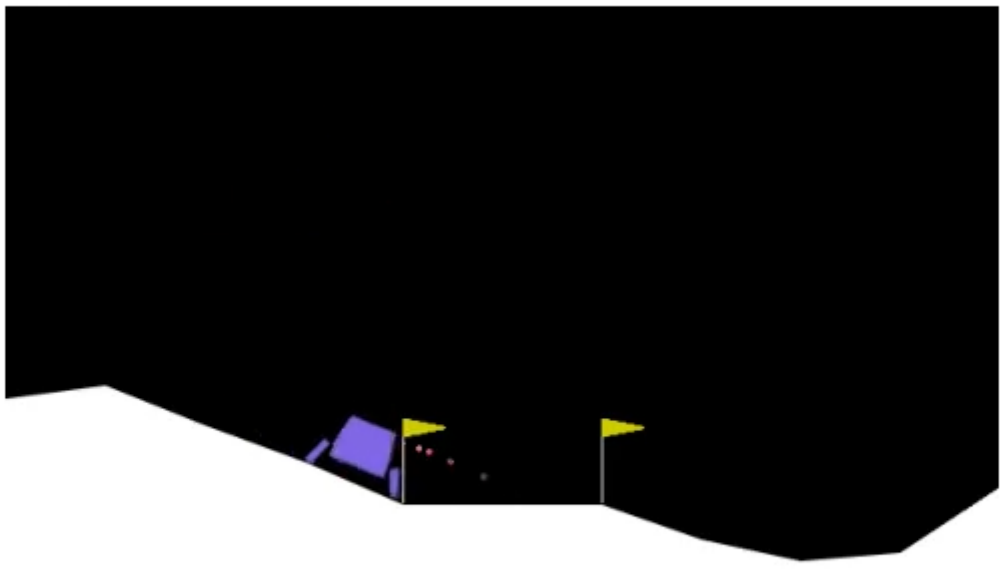
```
Episode 3400    Average Score: 230.62
Episode 3499    Average Score: 218.37
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3500    Average Score: 218.40
```



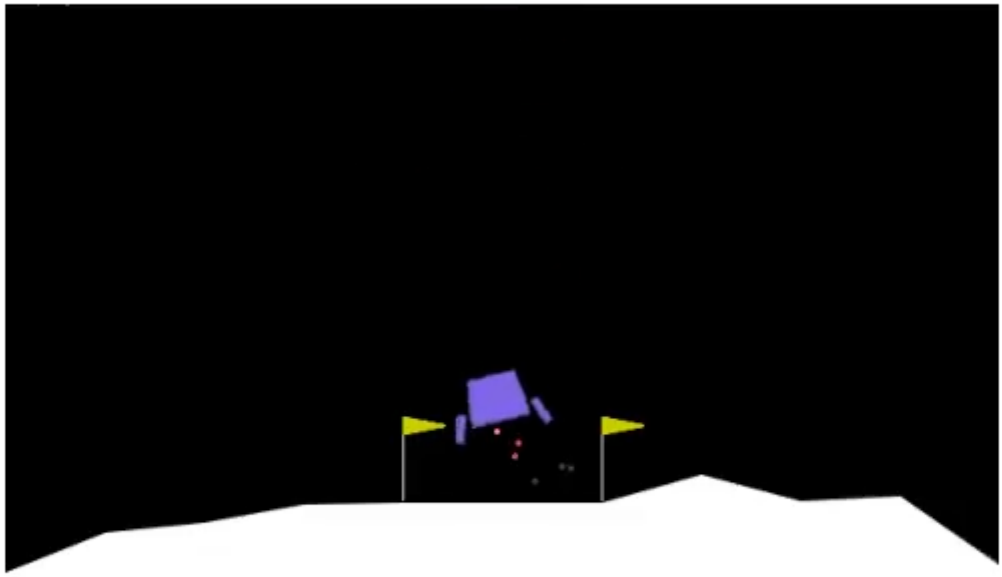
```
Episode 3500    Average Score: 218.40
Episode 3599    Average Score: 230.36
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3600    Average Score: 230.18
```



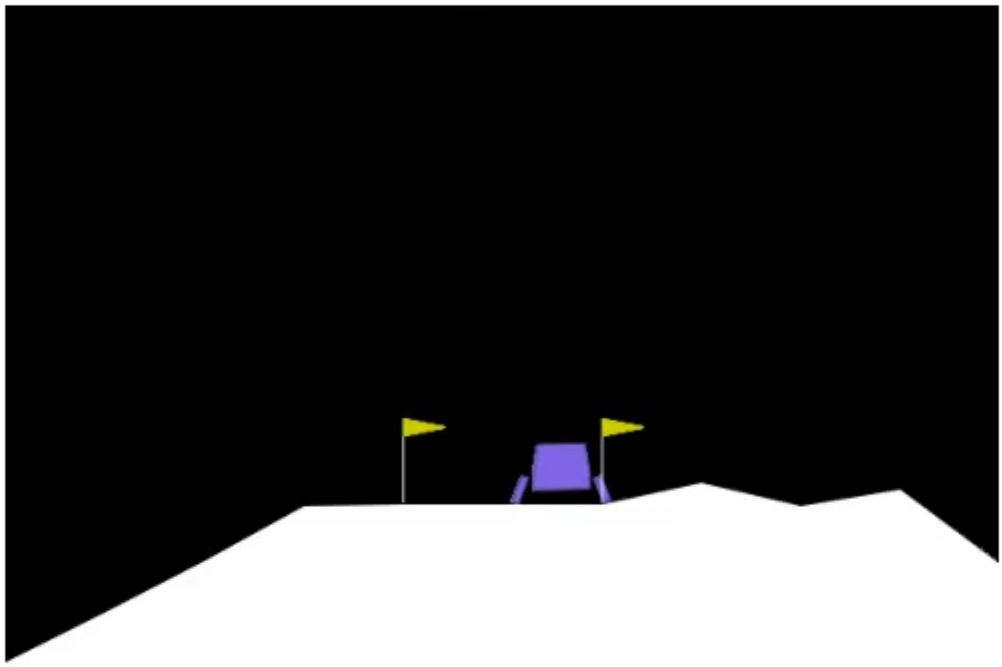
```
Episode 3600    Average Score: 230.18
Episode 3699    Average Score: 231.66
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3700    Average Score: 232.16
```



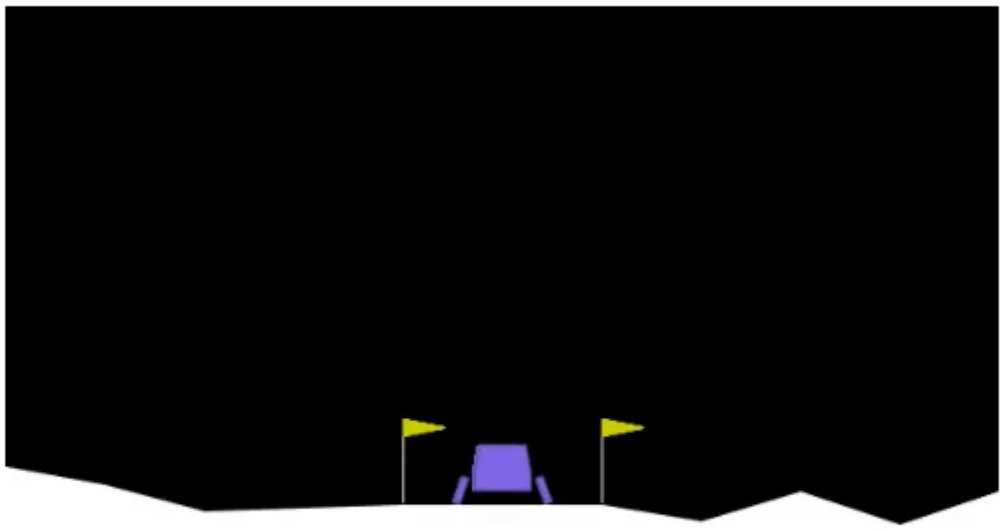
```
Episode 3700    Average Score: 232.16
Episode 3799    Average Score: 238.68
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3800    Average Score: 235.16
```



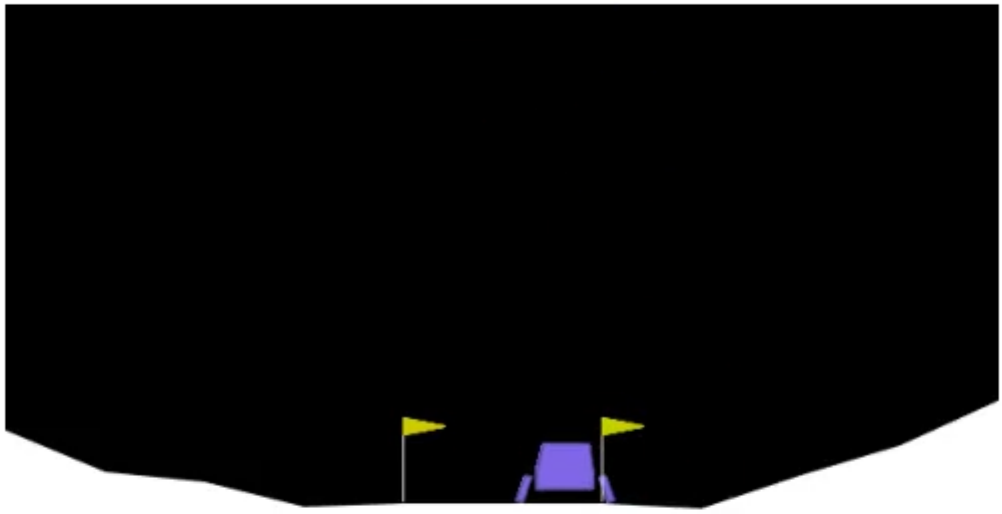
```
Episode 3800    Average Score: 235.16
Episode 3899    Average Score: 239.95
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 3900    Average Score: 243.40
```



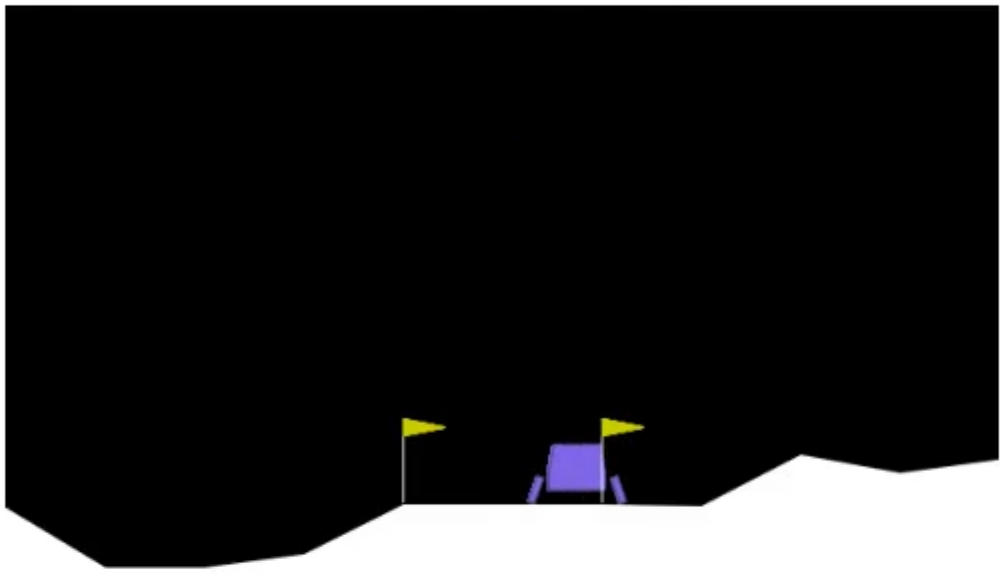
```
Episode 3900    Average Score: 243.40
Episode 3999    Average Score: 254.86
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4000    Average Score: 254.81
```



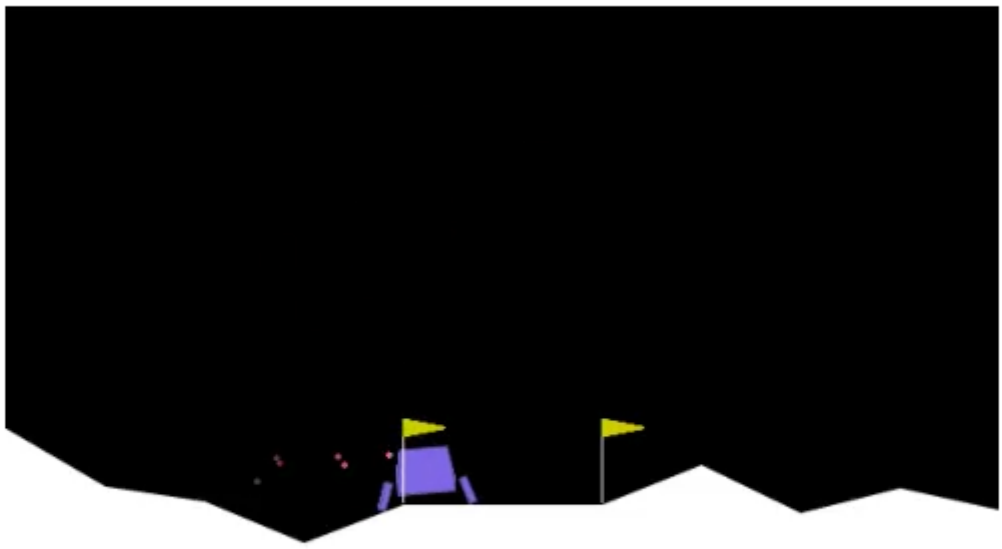
```
Episode 4000    Average Score: 254.81
Episode 4099    Average Score: 265.20
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4100    Average Score: 265.35
```



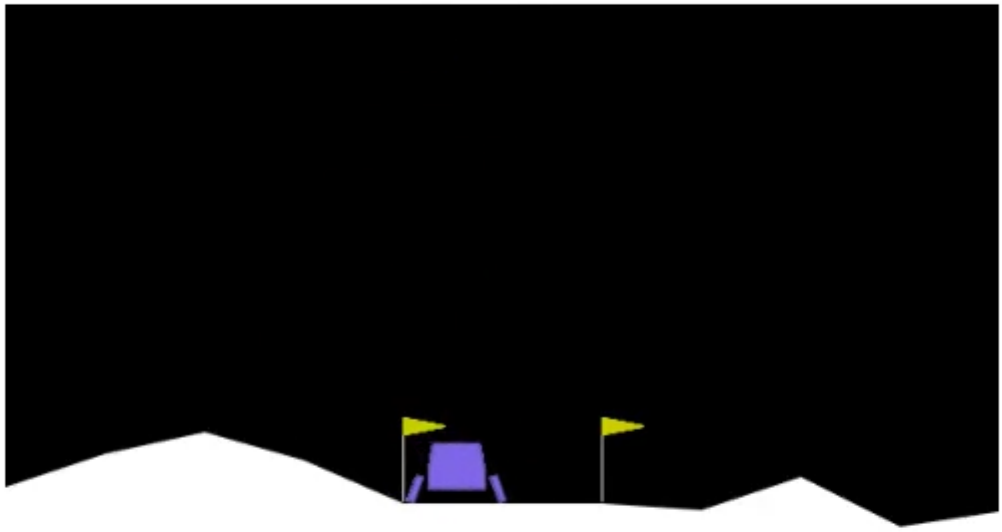
```
Episode 4100    Average Score: 265.35
Episode 4199    Average Score: 265.37
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4200    Average Score: 265.29
```



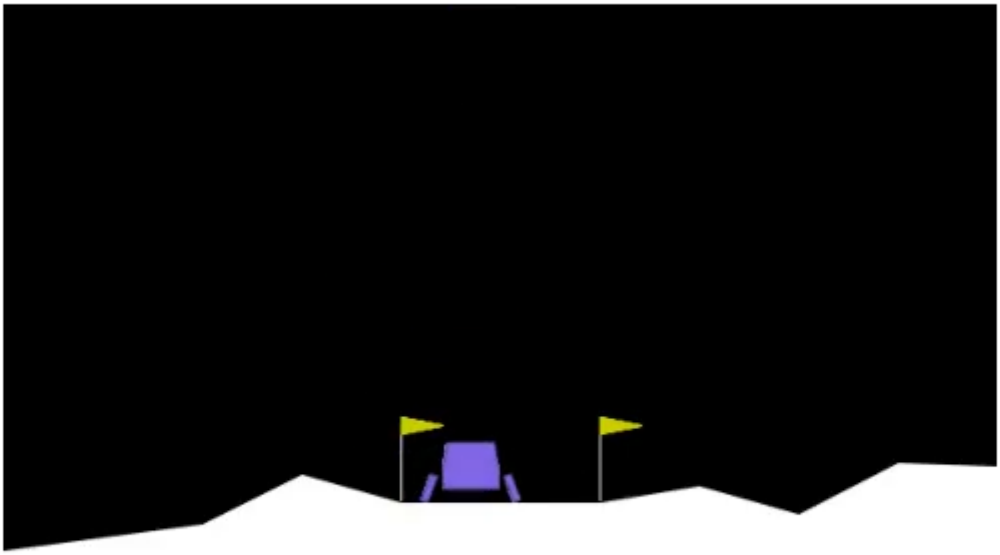
```
Episode 4200    Average Score: 265.29
Episode 4299    Average Score: 267.79
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4300    Average Score: 267.62
```



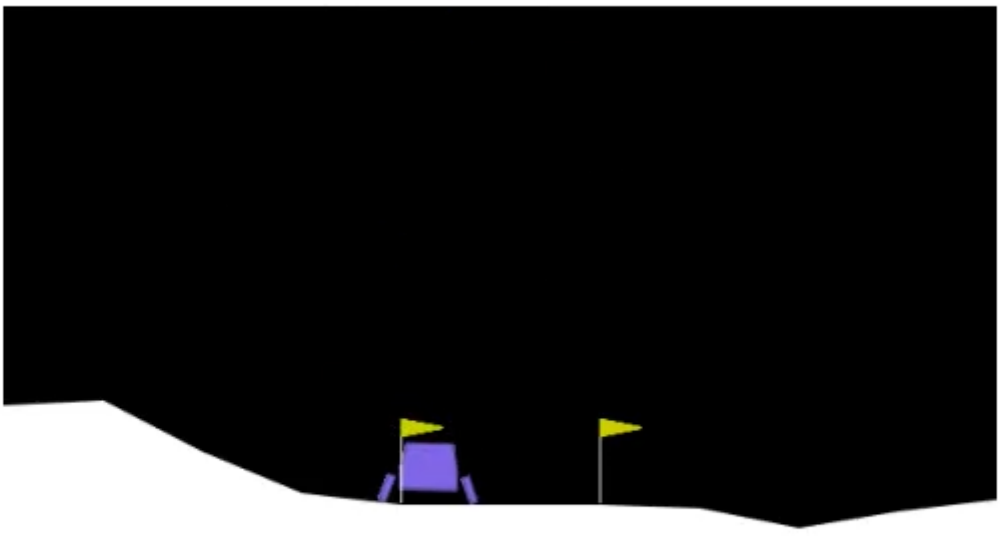
```
Episode 4300    Average Score: 267.62
Episode 4399    Average Score: 269.58
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4400    Average Score: 270.00
```



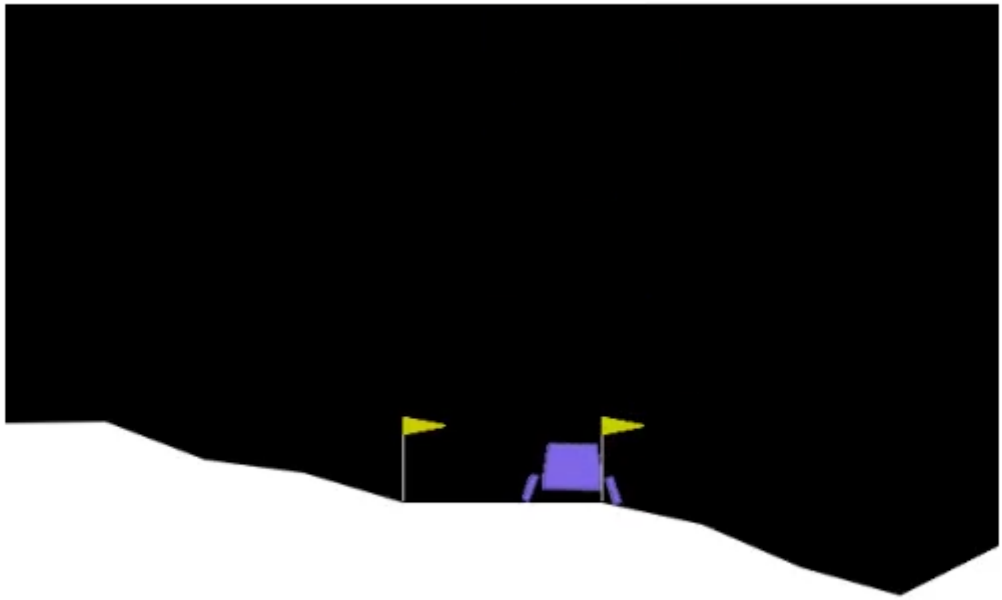
```
Episode 4400    Average Score: 270.00
Episode 4499    Average Score: 273.90
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4500    Average Score: 273.56
```



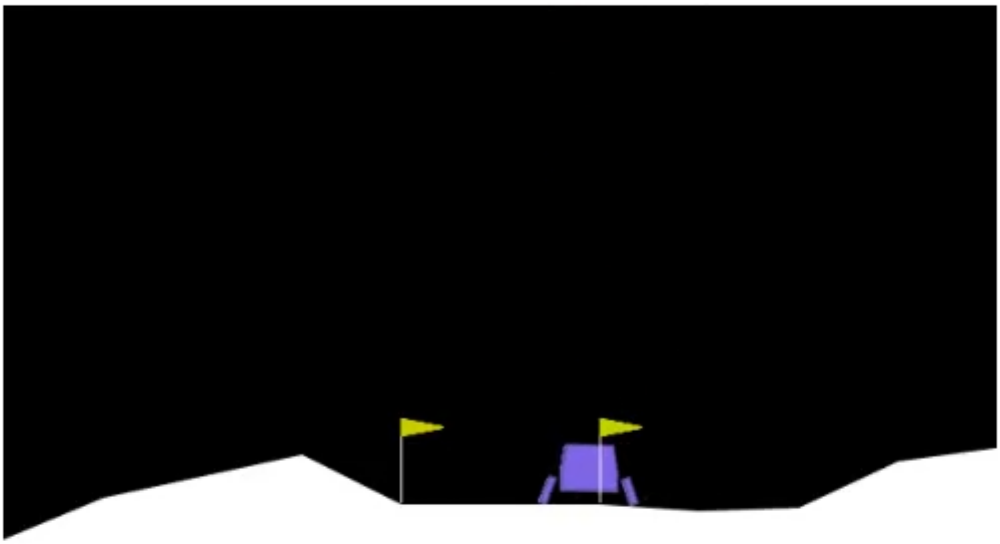
```
Episode 4500    Average Score: 273.56
Episode 4599    Average Score: 258.12
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4600    Average Score: 258.30
```



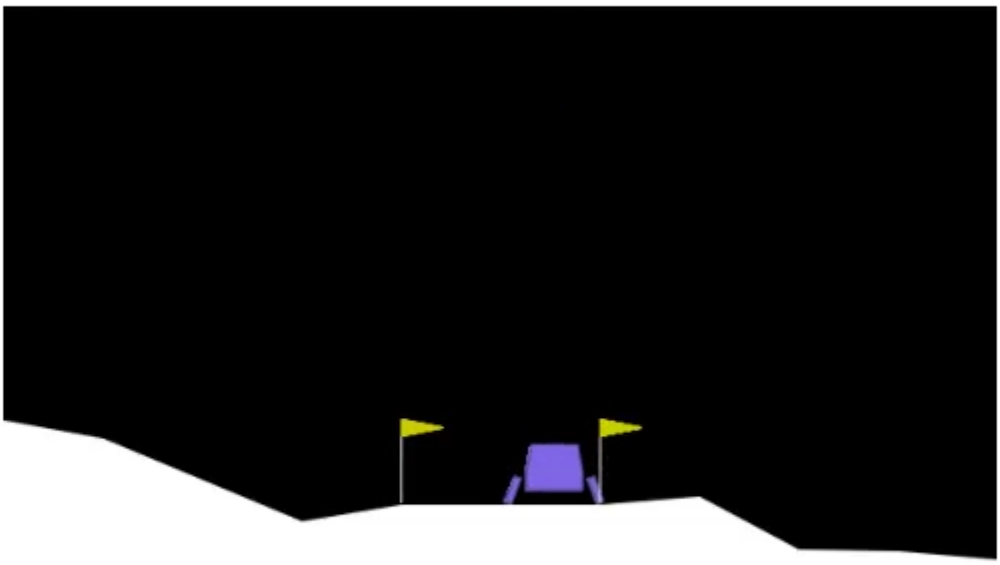
```
Episode 4600    Average Score: 258.30
Episode 4699    Average Score: 265.81
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4700    Average Score: 265.66
```



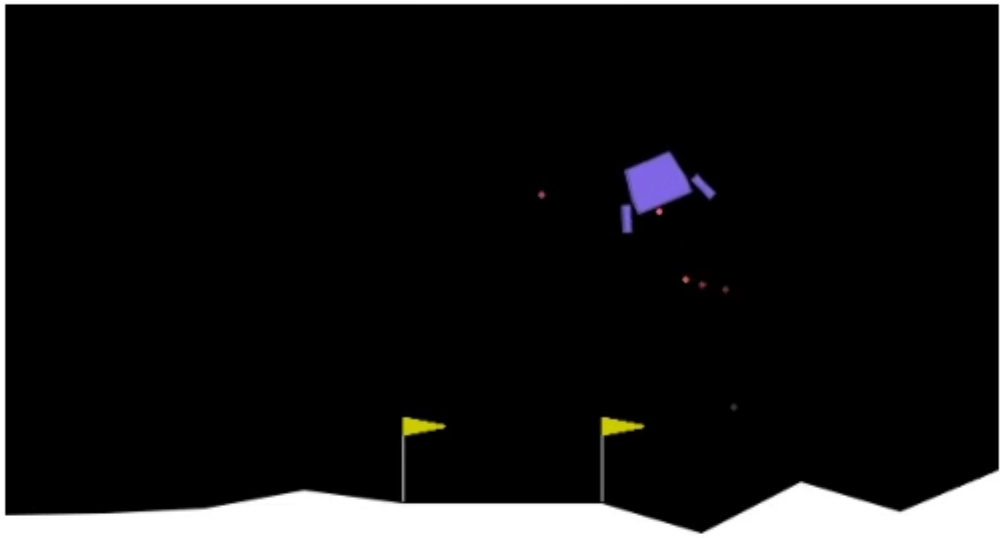
```
Episode 4700    Average Score: 265.66
Episode 4799    Average Score: 273.14
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4800    Average Score: 273.15
```



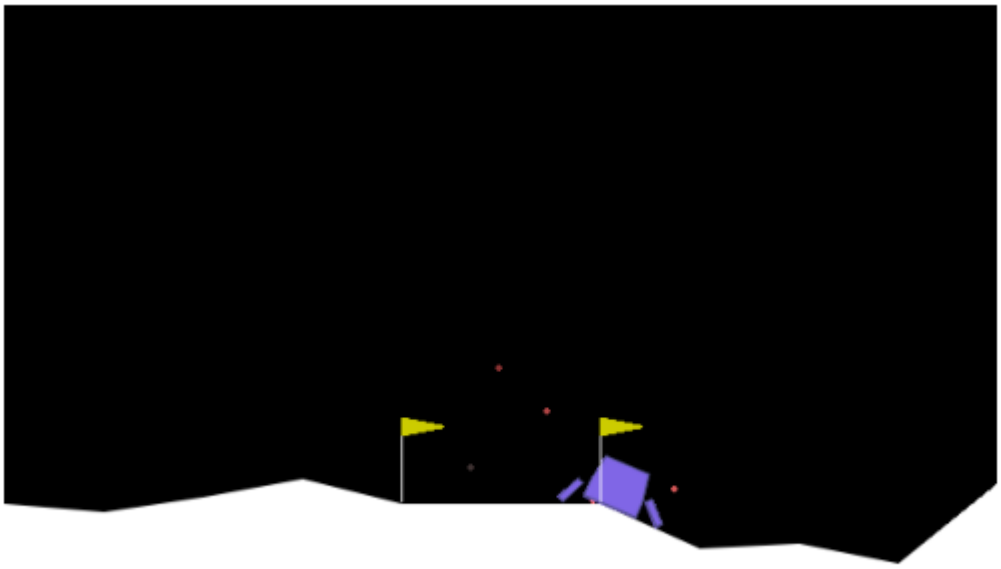
```
Episode 4800    Average Score: 273.15
Episode 4899    Average Score: 272.26
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 4900    Average Score: 272.36
```

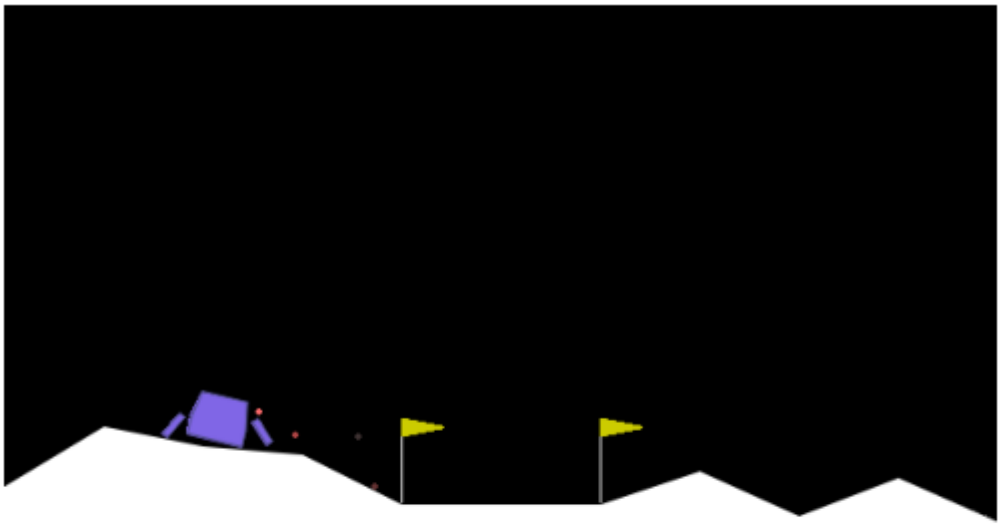


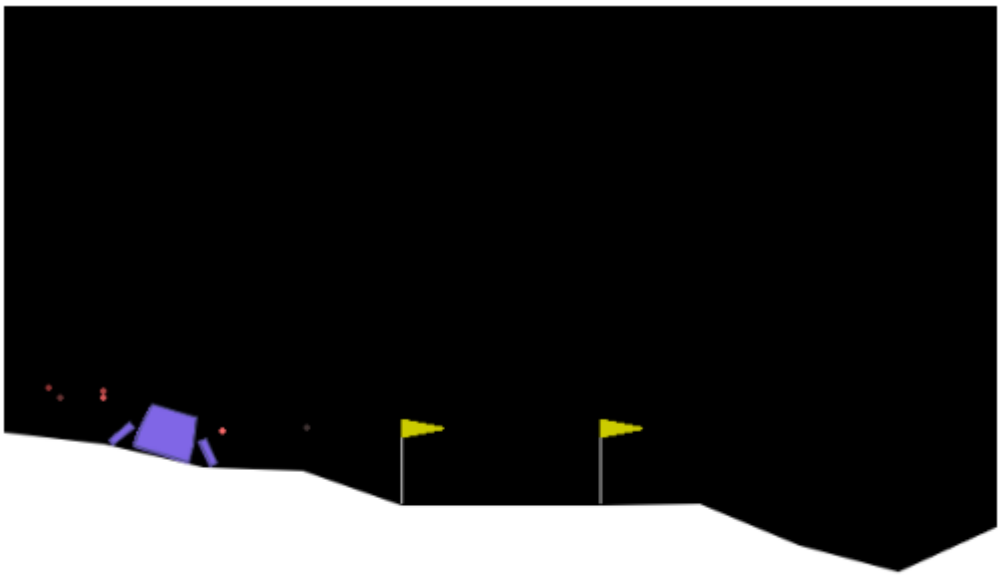
```
Episode 4900    Average Score: 272.36
Episode 4999    Average Score: 277.54
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Episode 5000    Average Score: 276.50
```

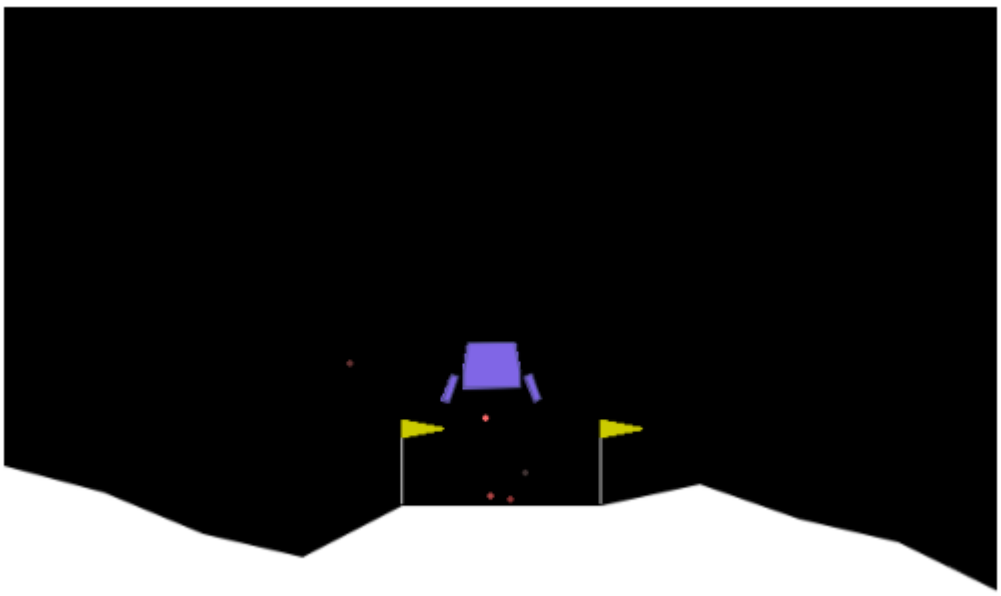


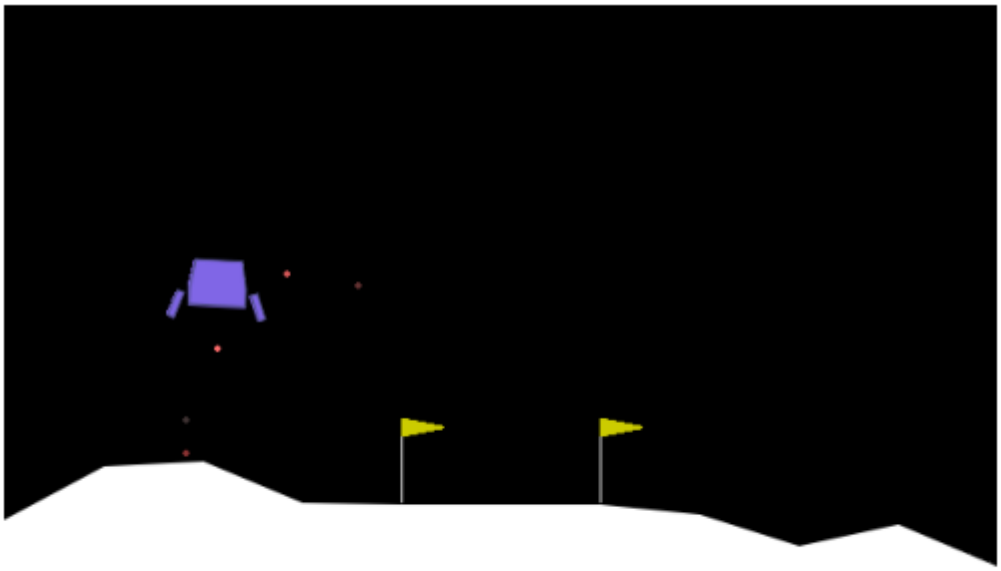
Episode 5000   Average Score: 276.50

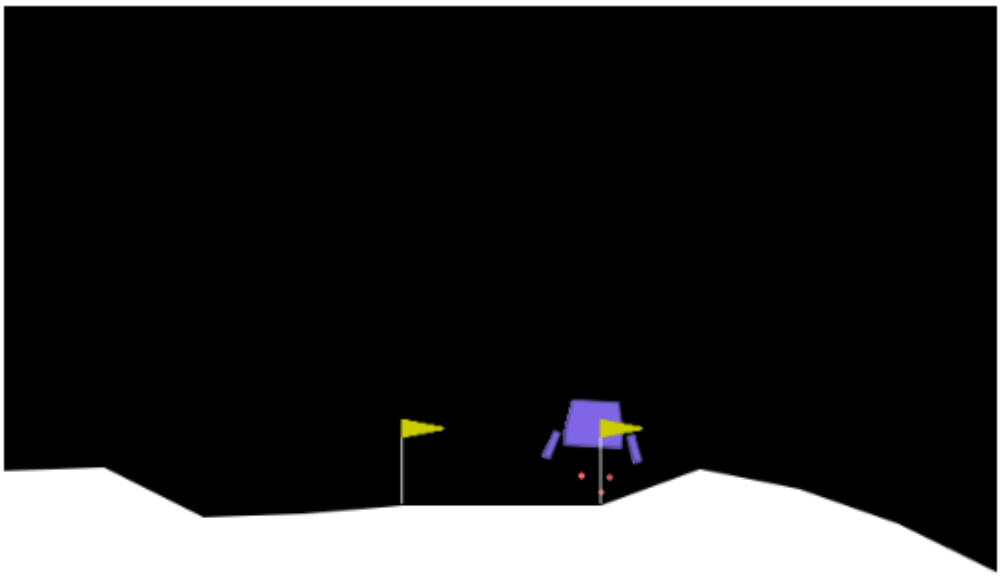


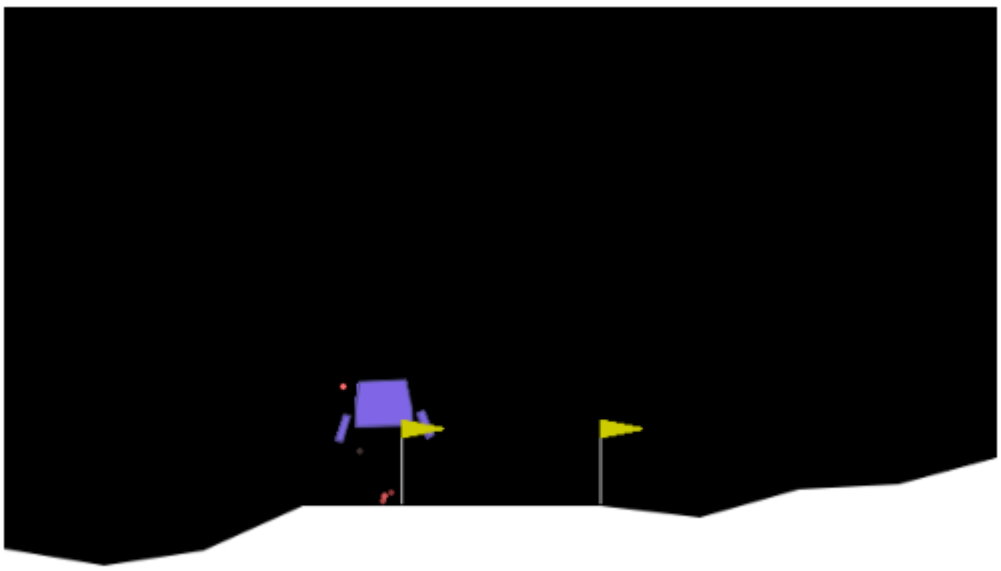


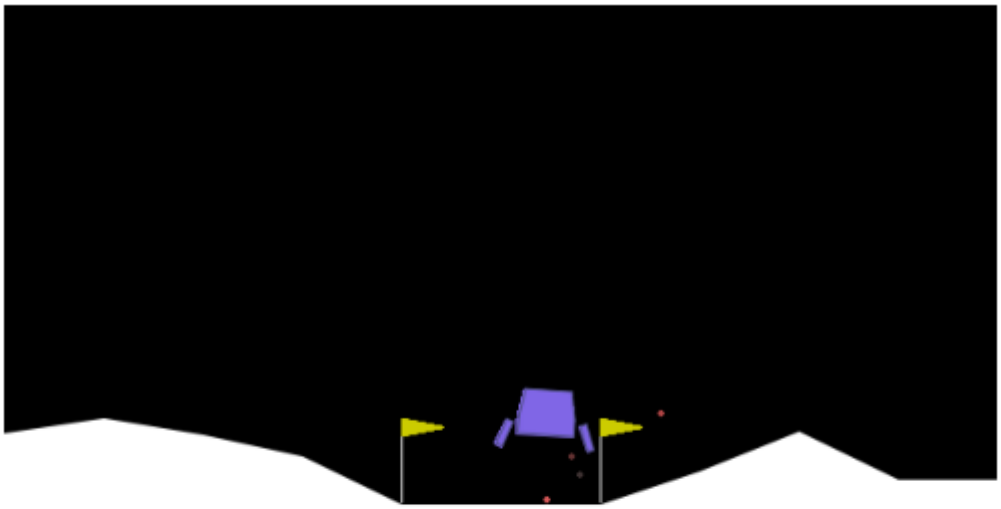


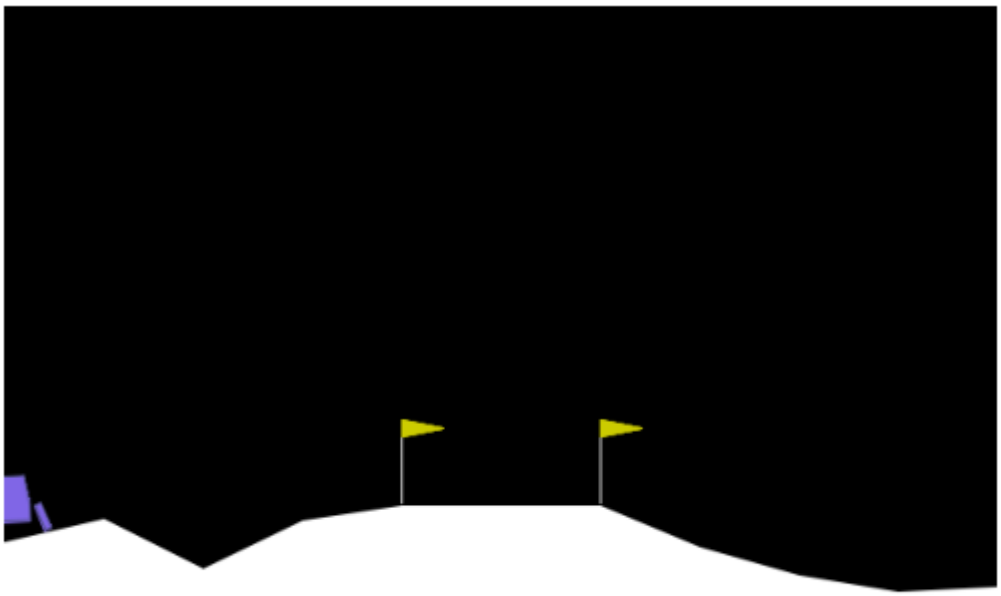


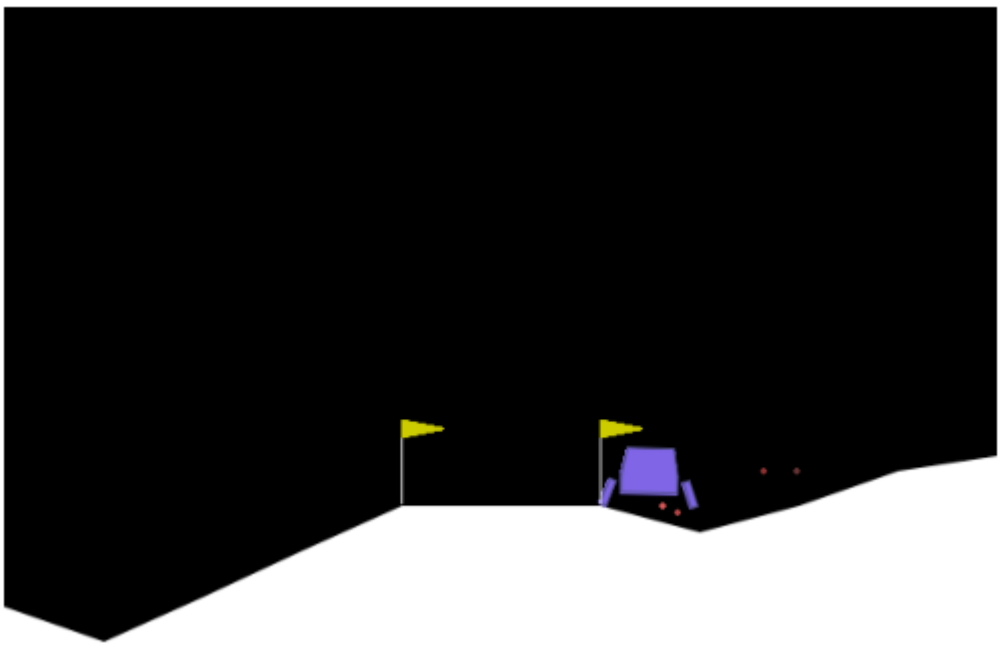


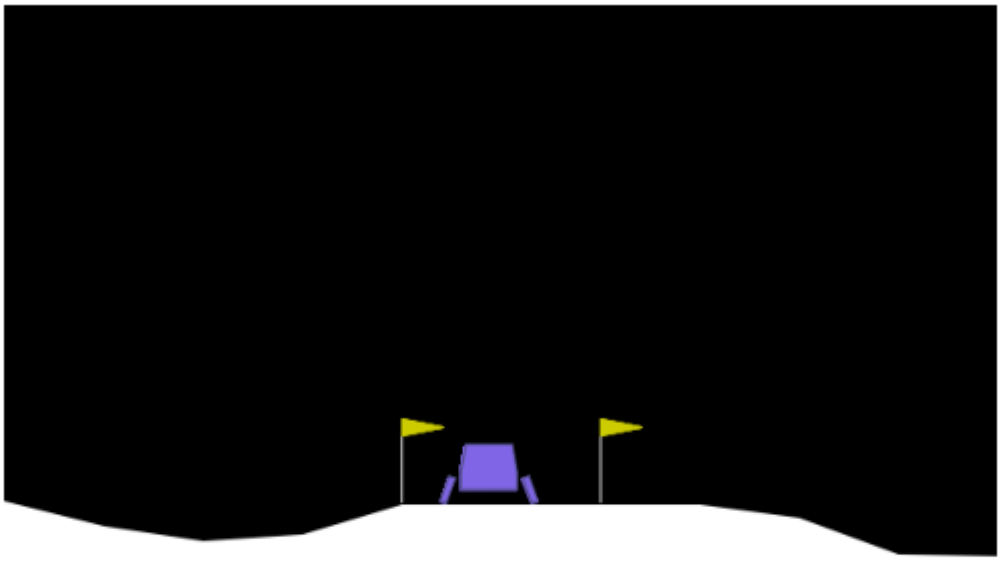


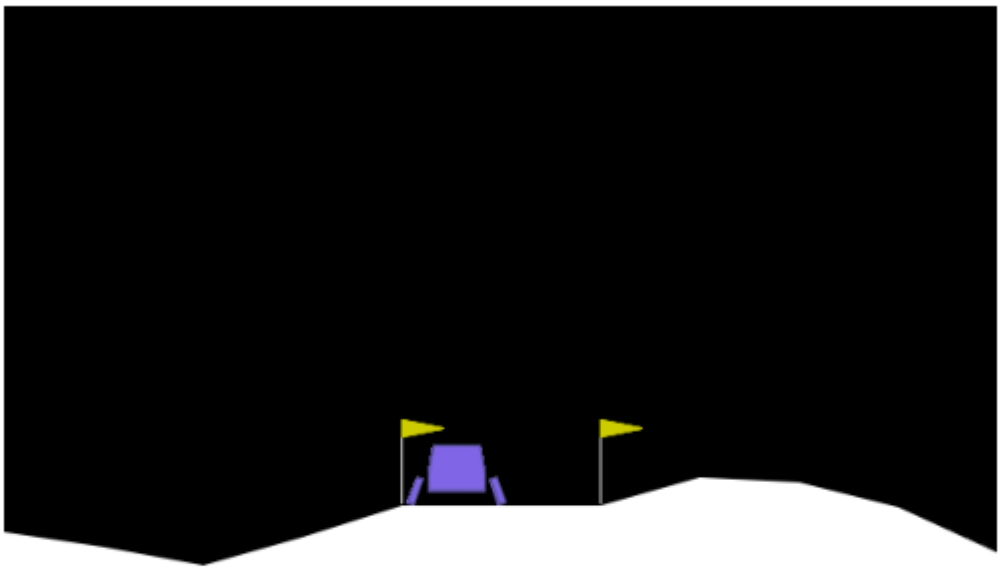


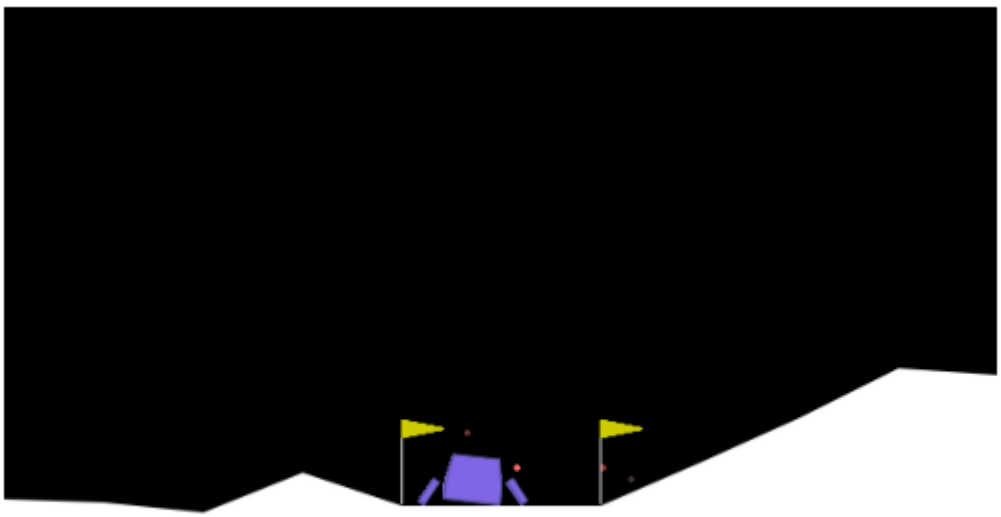


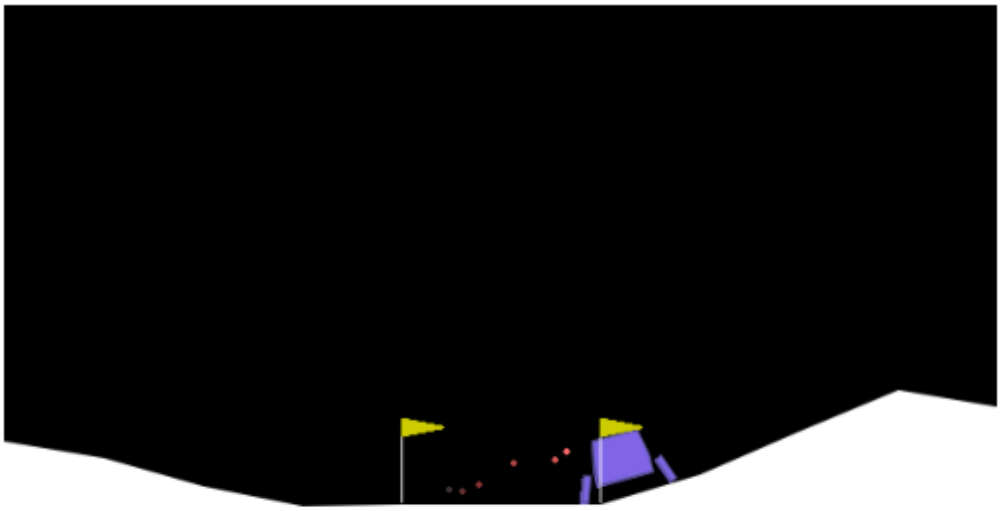


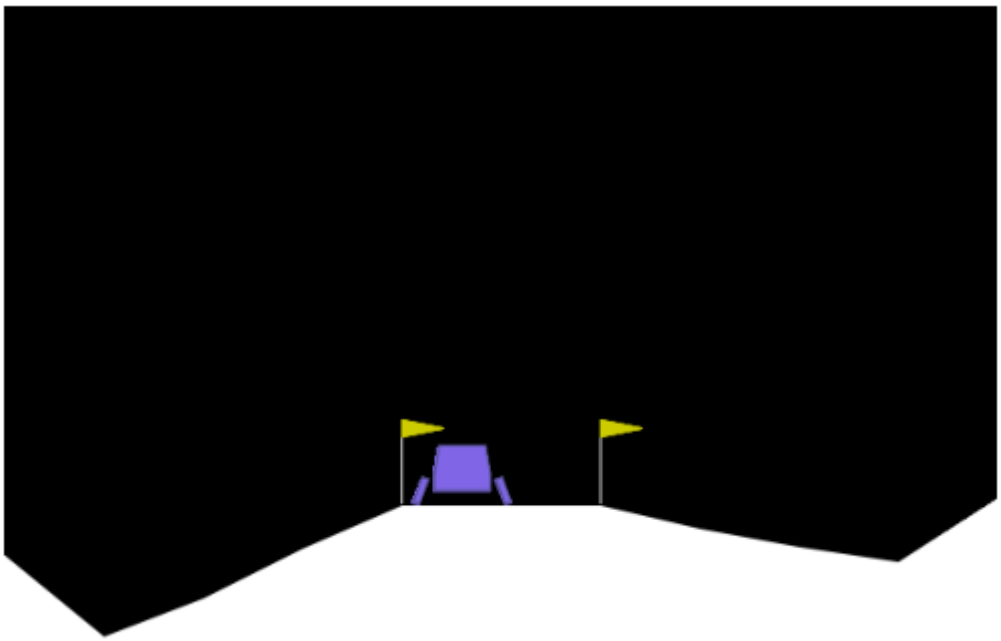


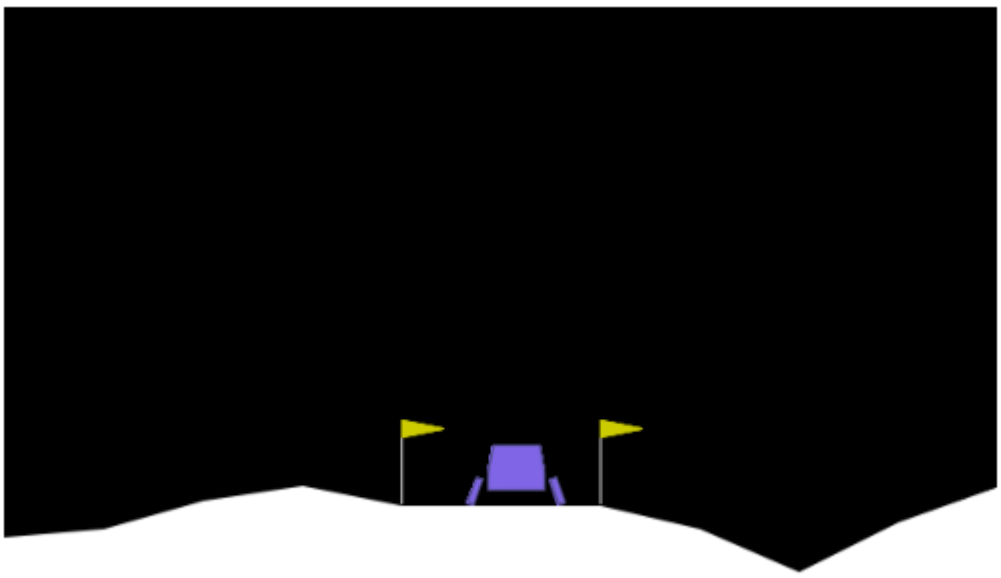


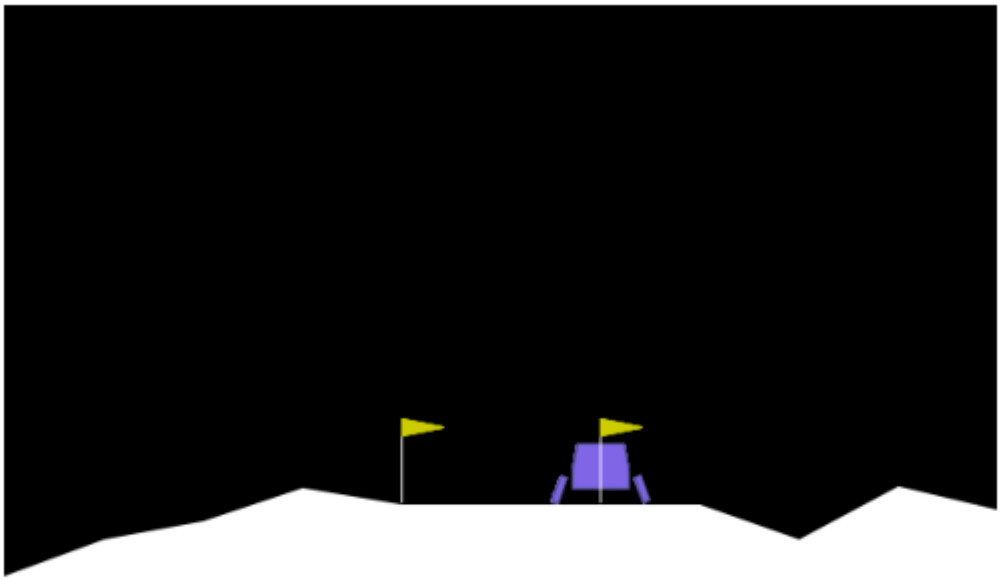


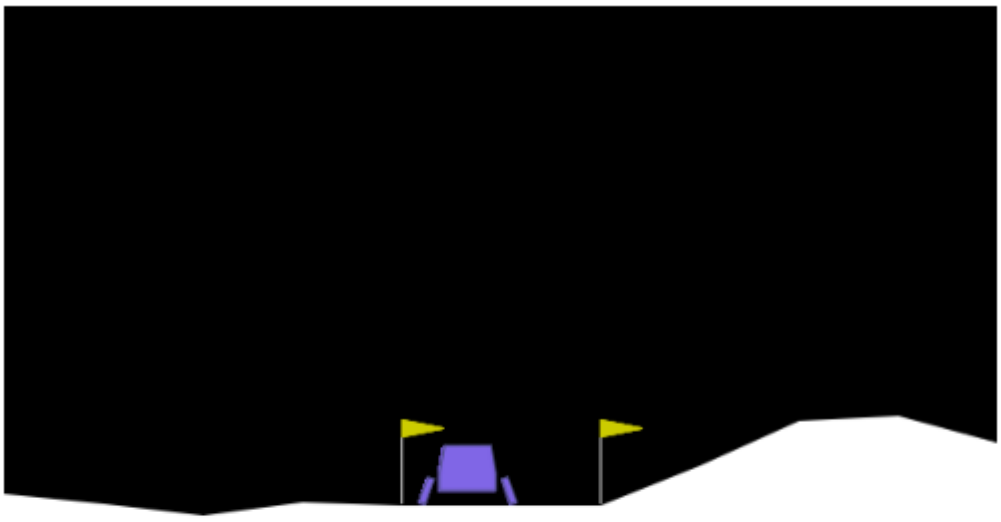


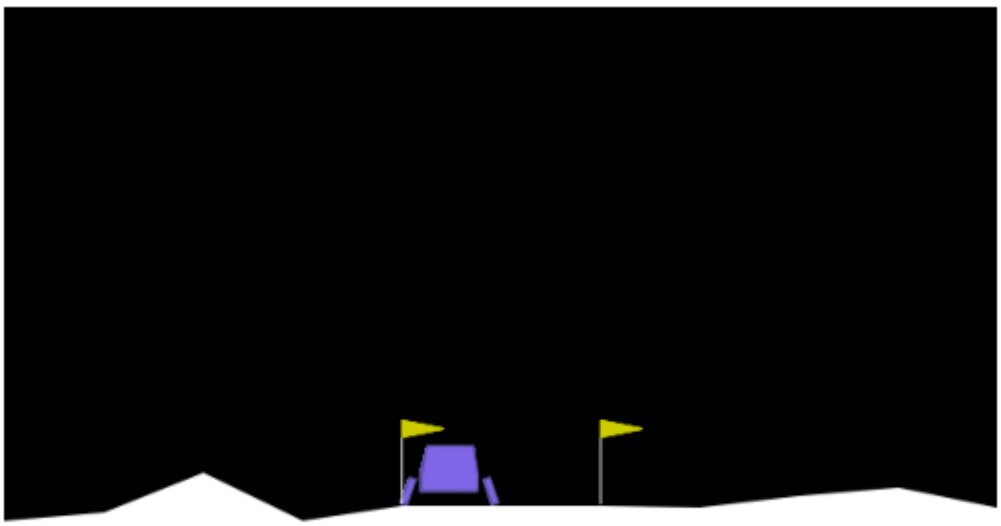


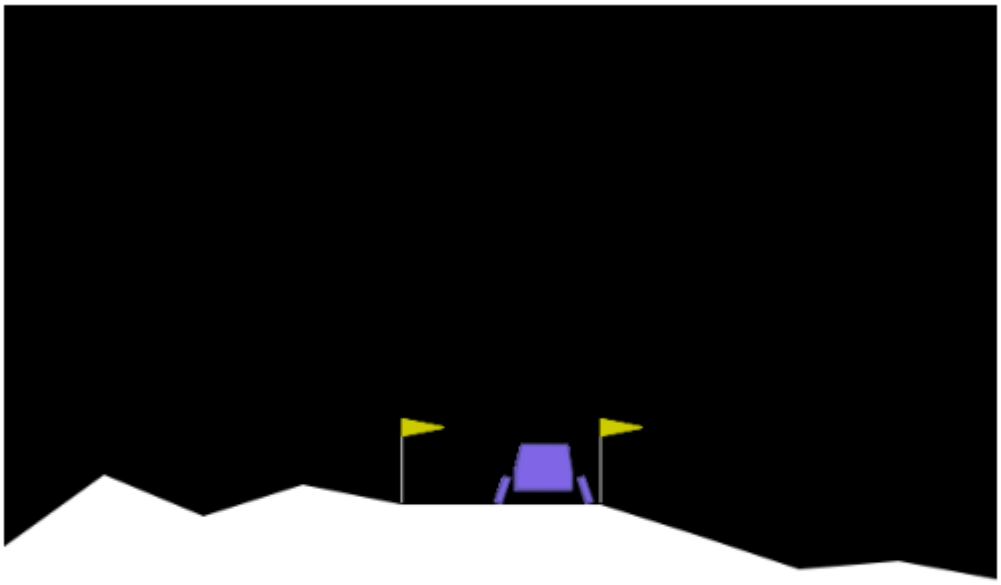


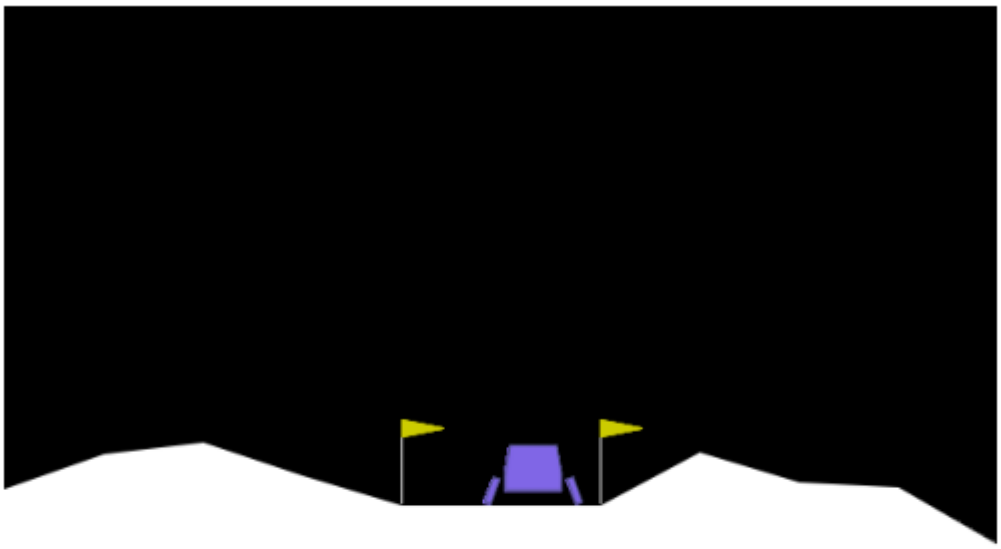


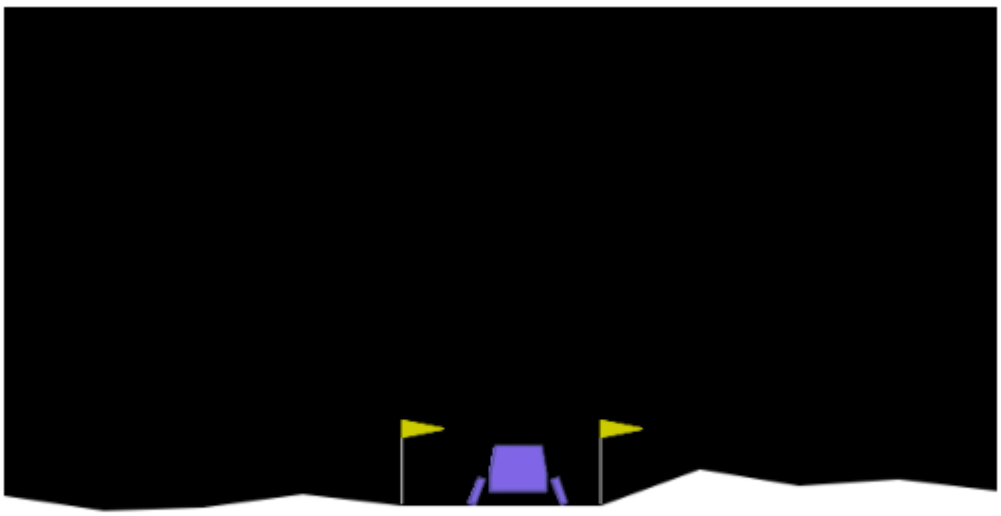


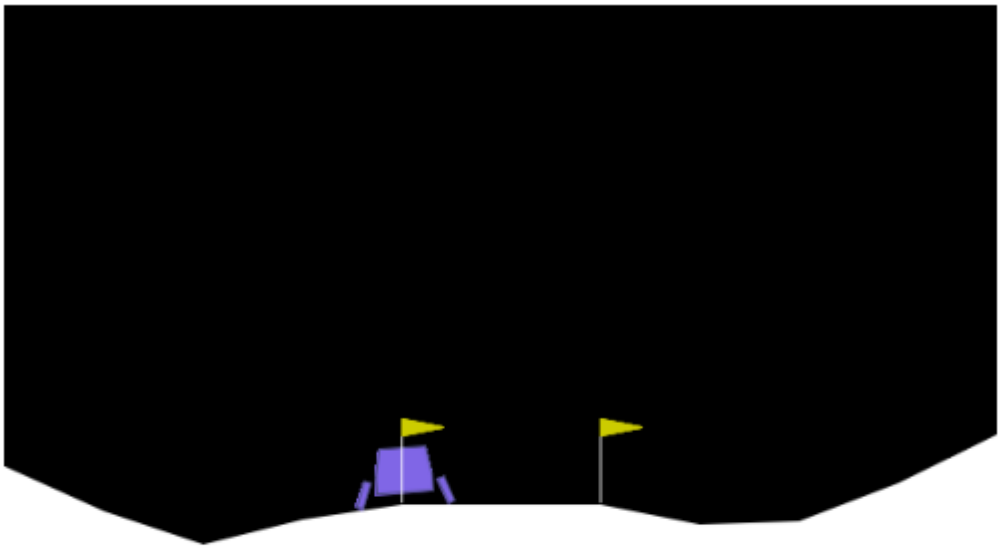


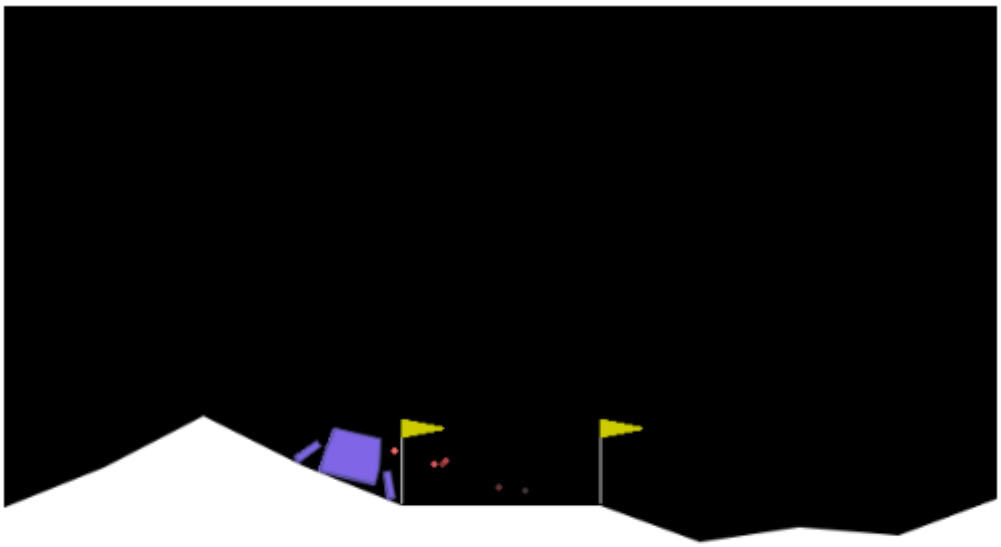


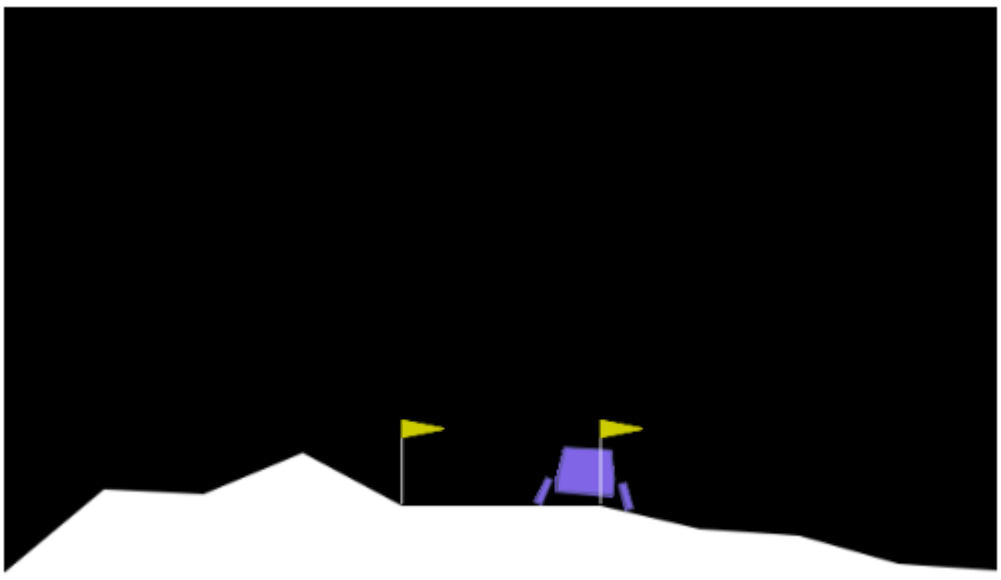


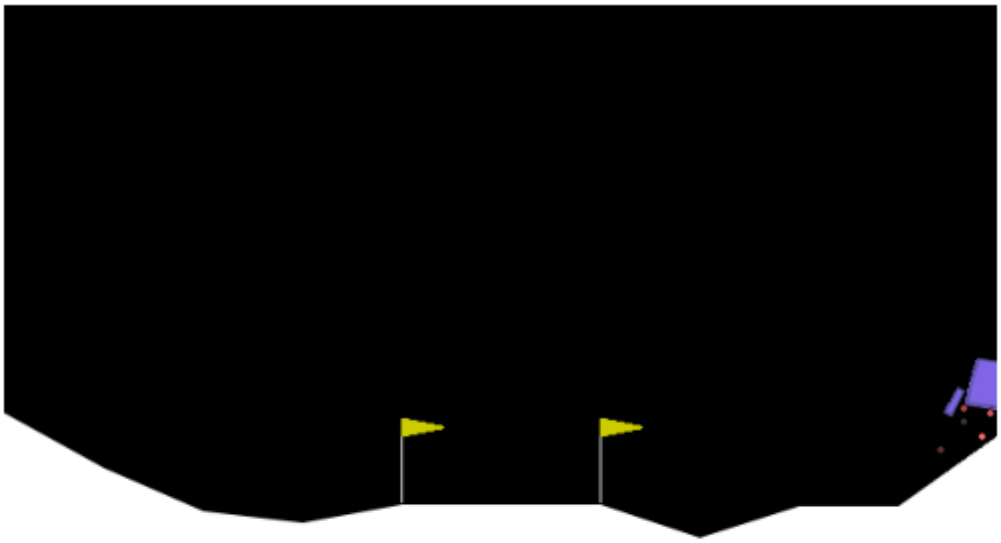


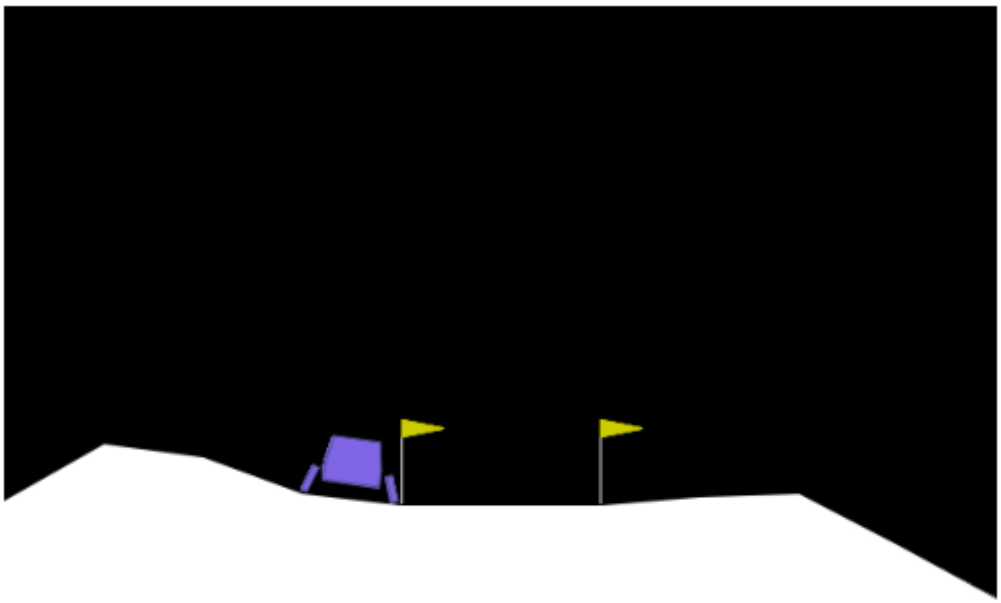


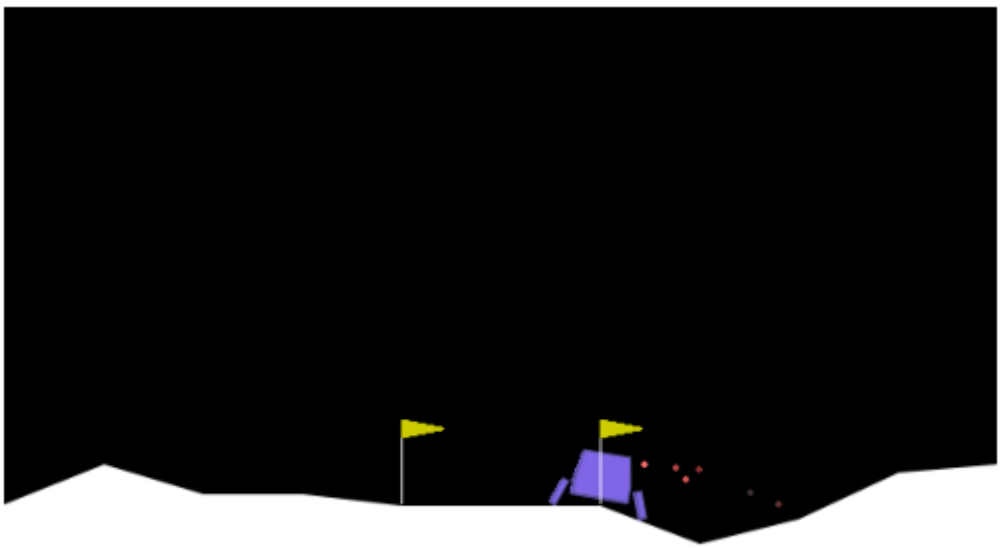


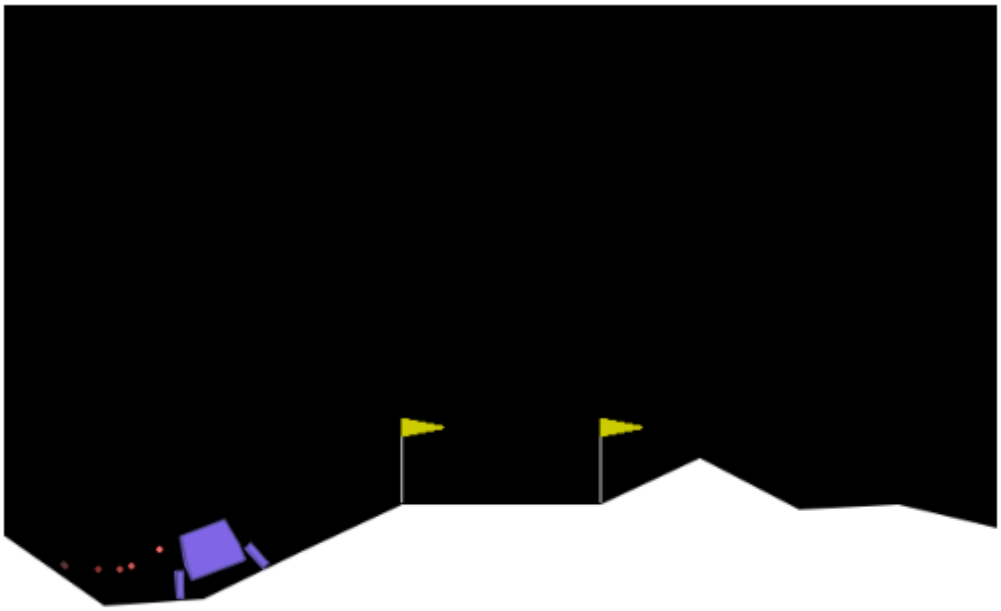


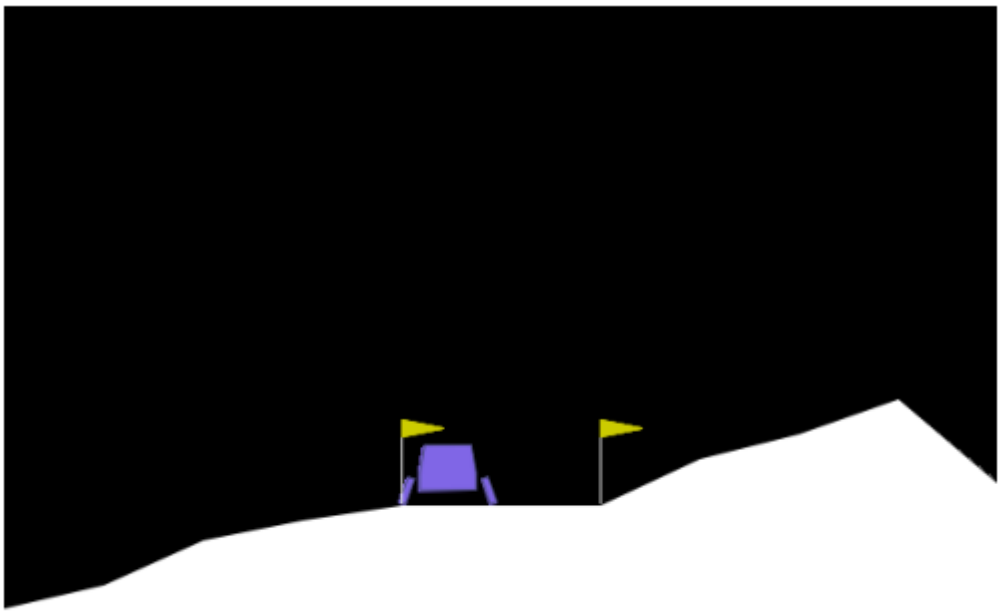


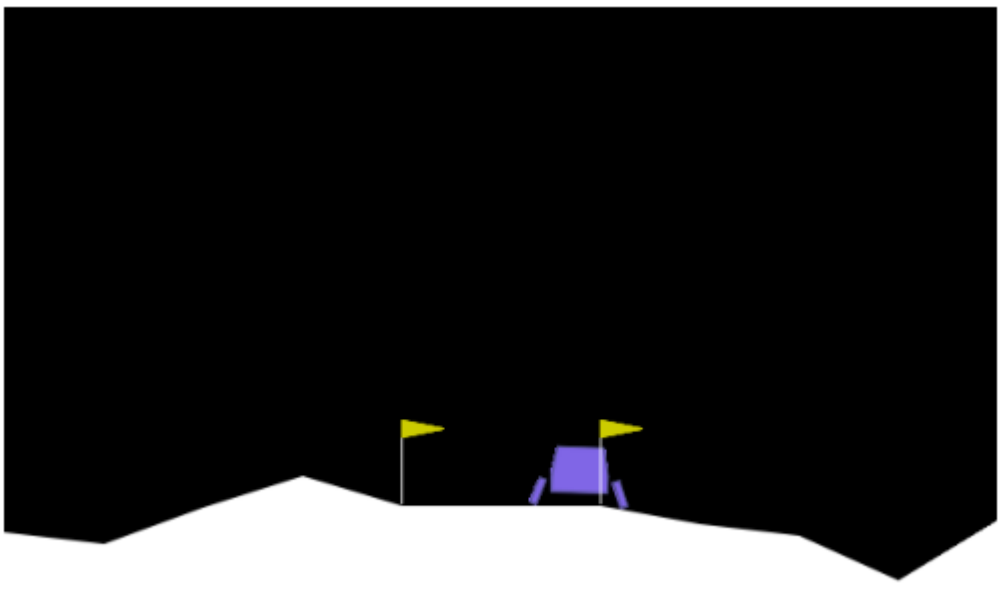


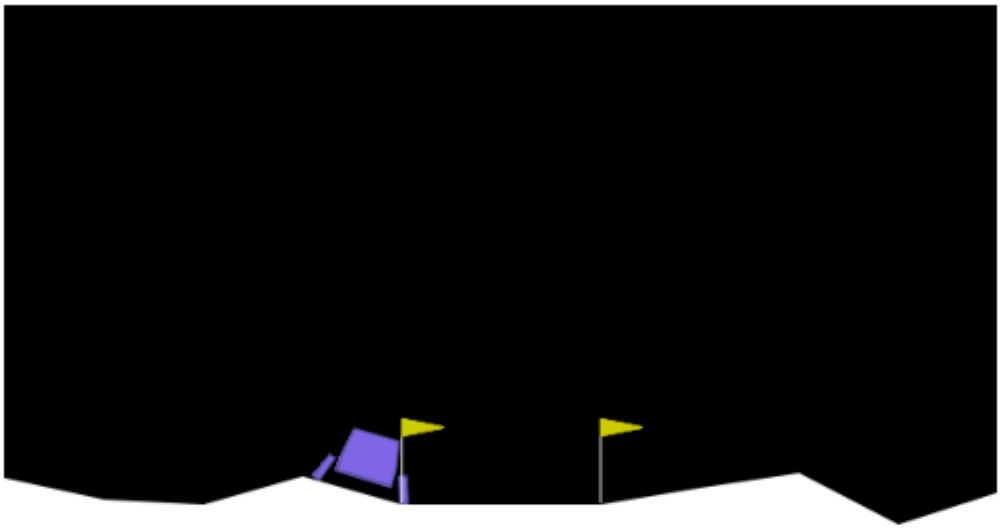


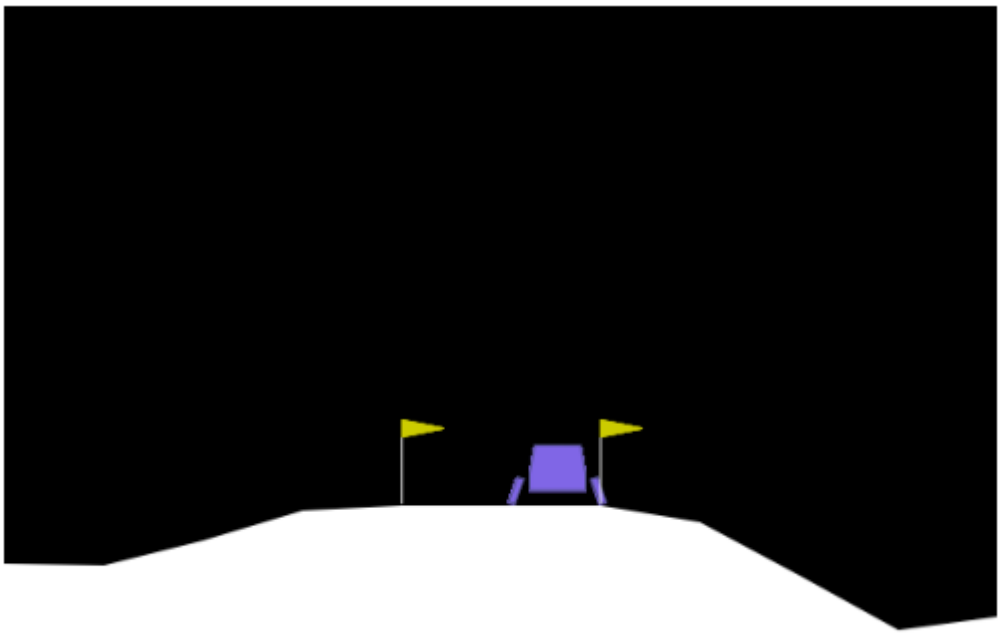


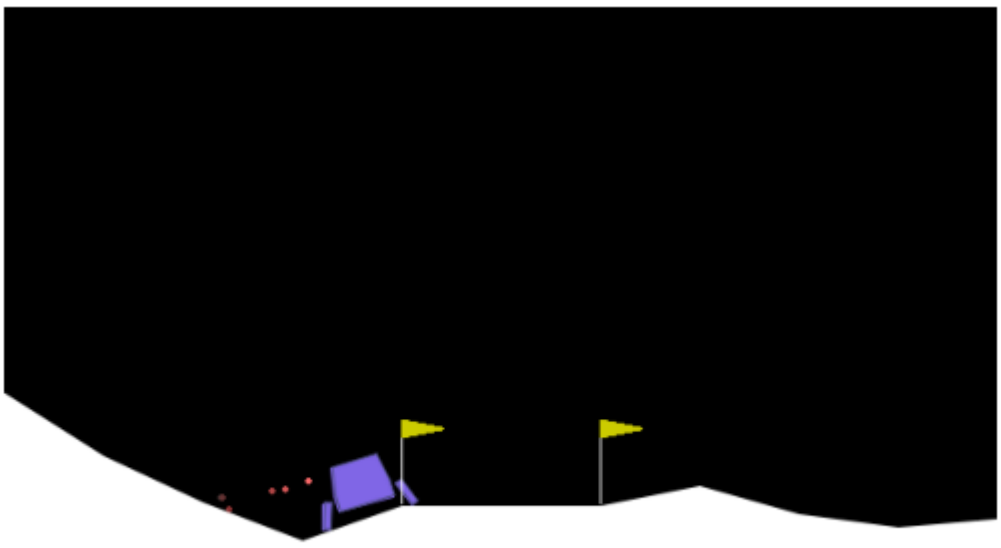


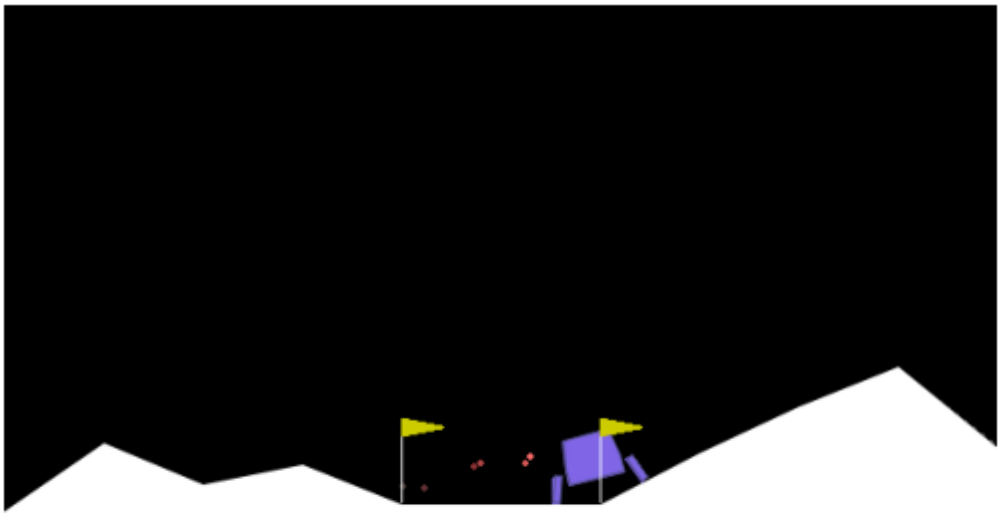


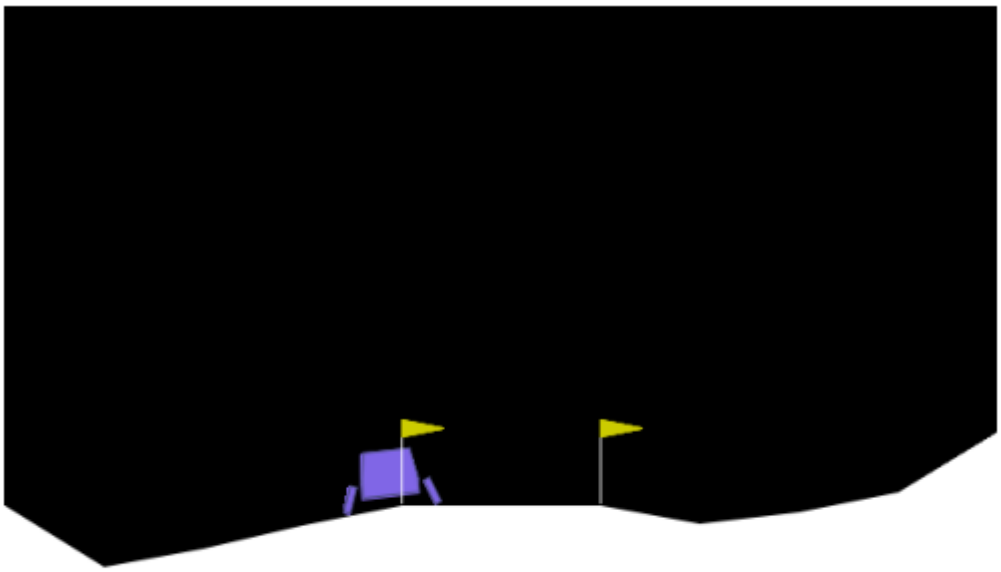


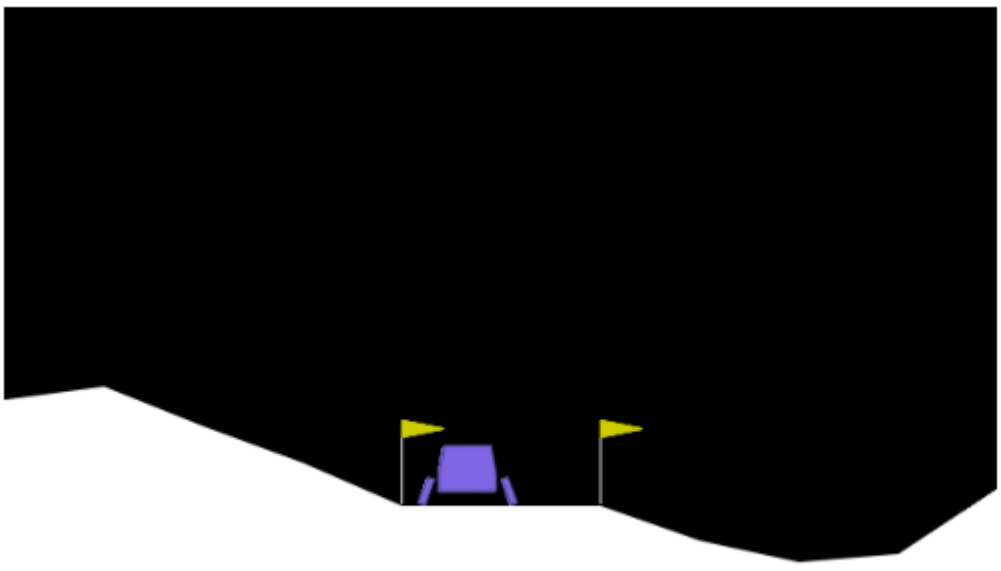


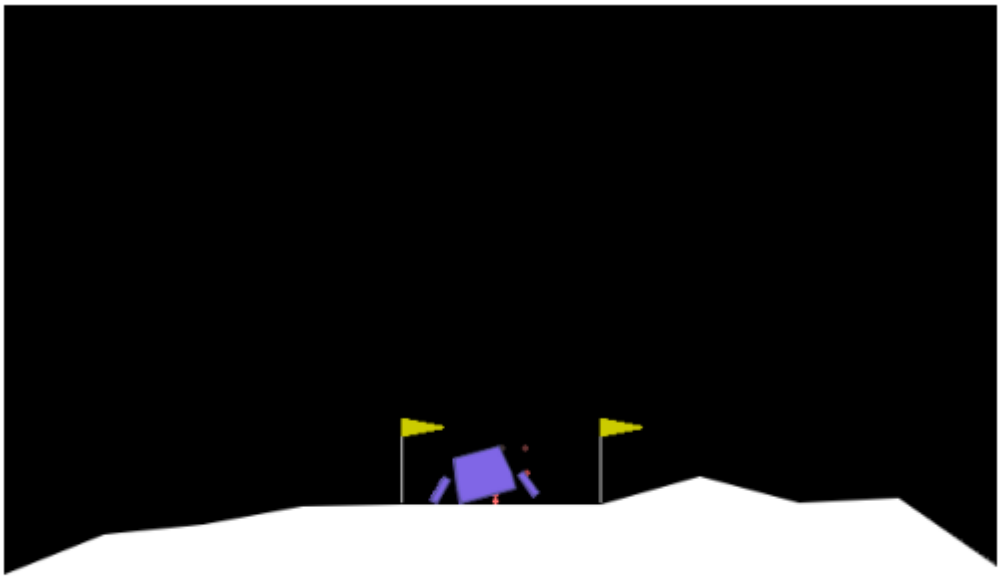


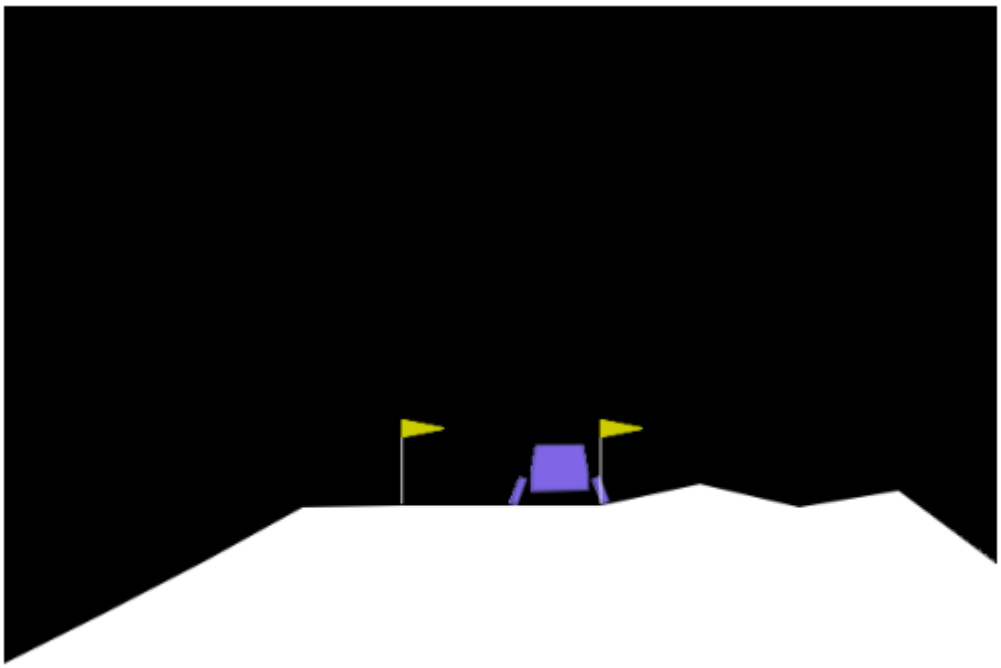


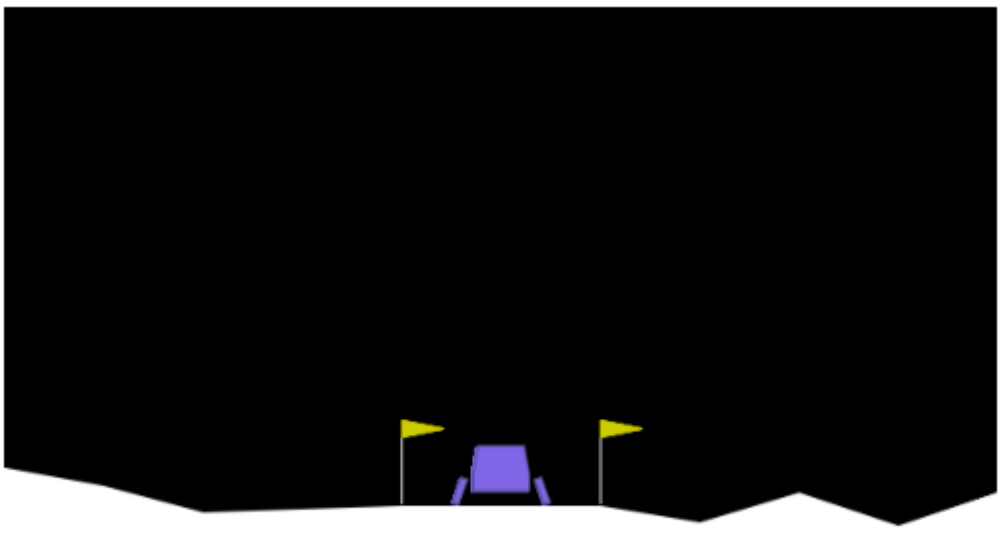


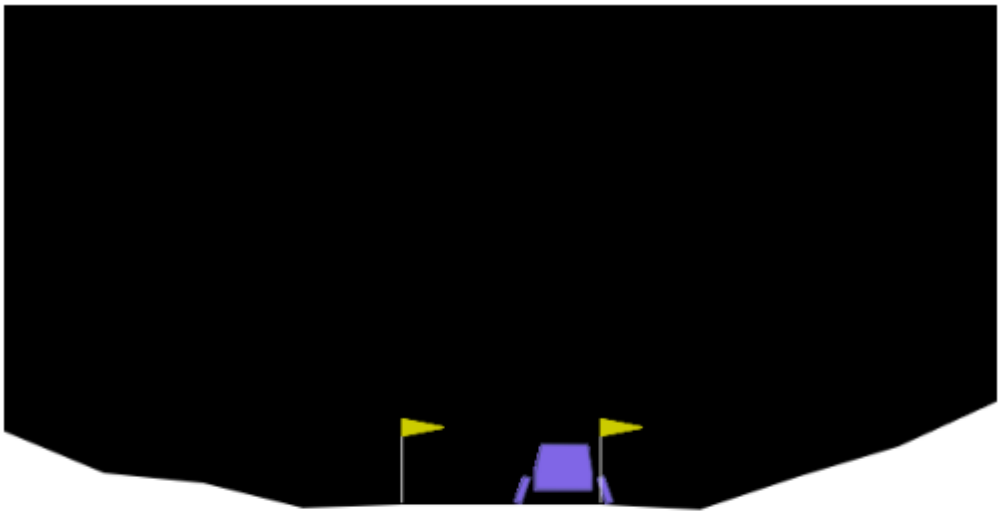


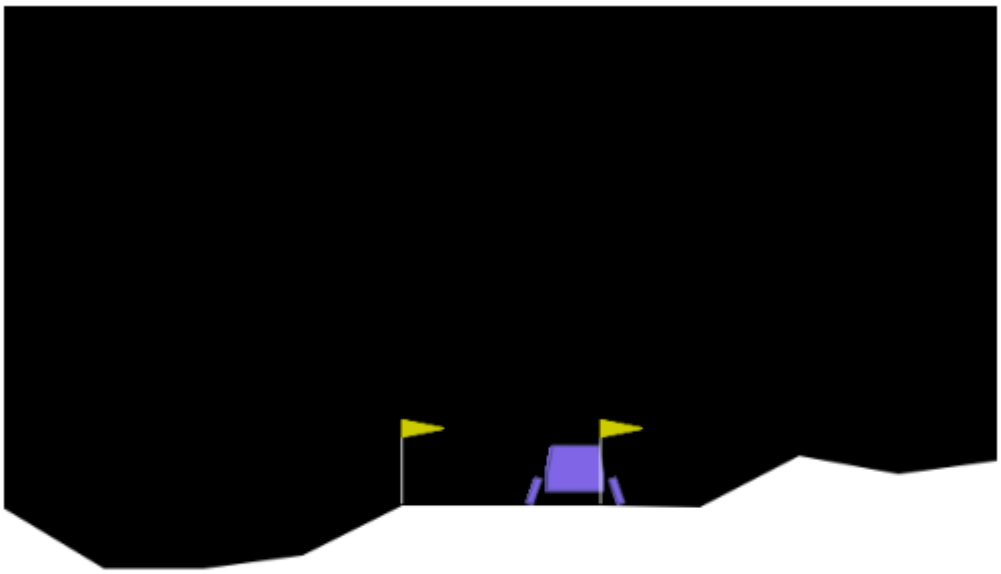


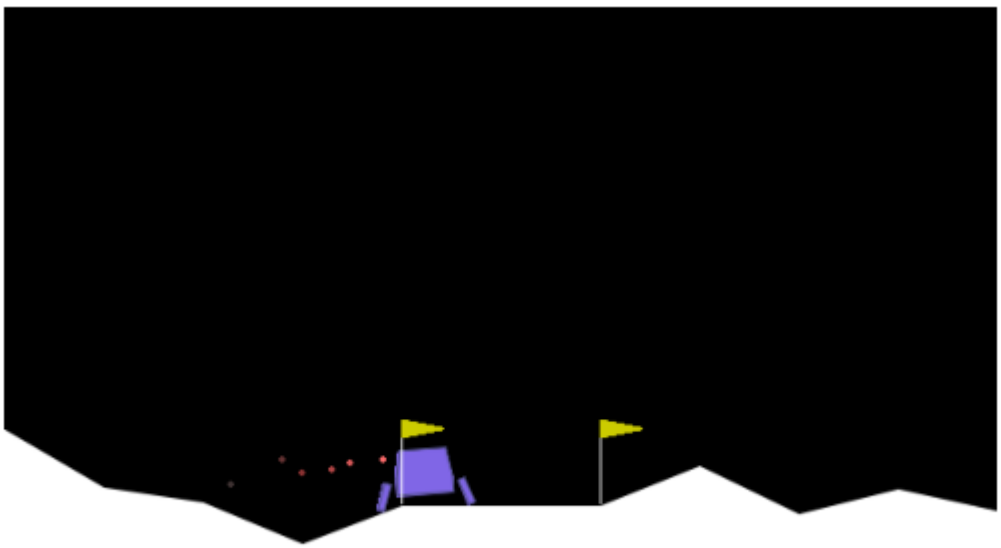


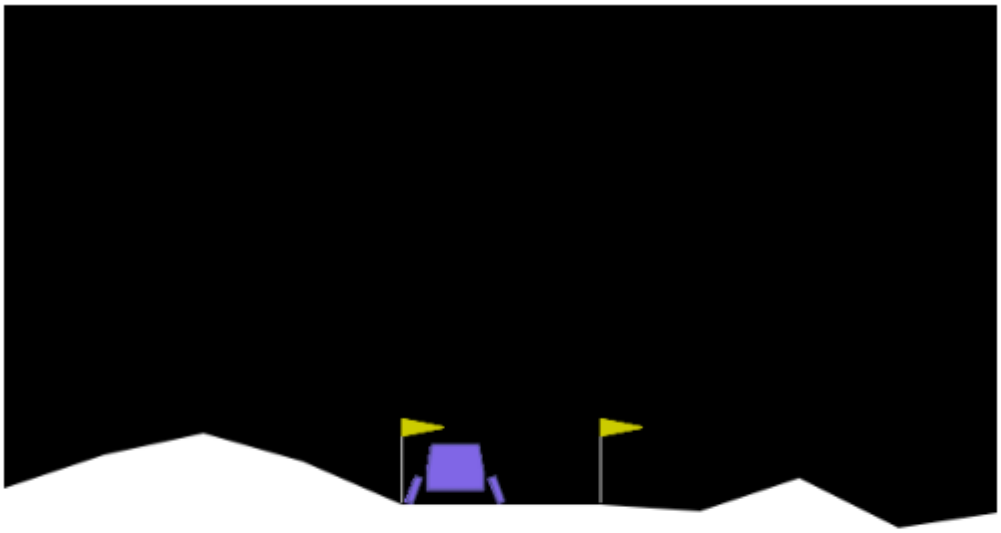


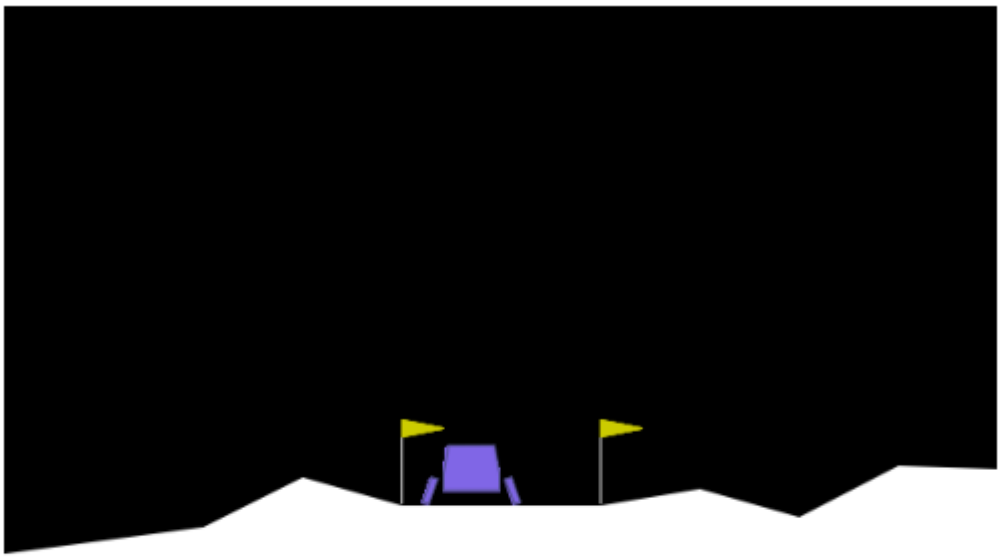


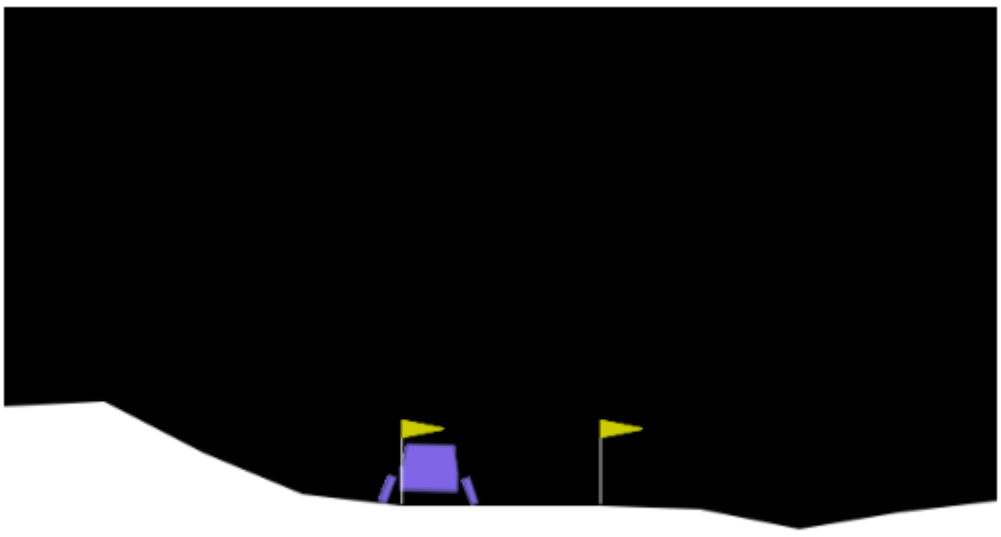


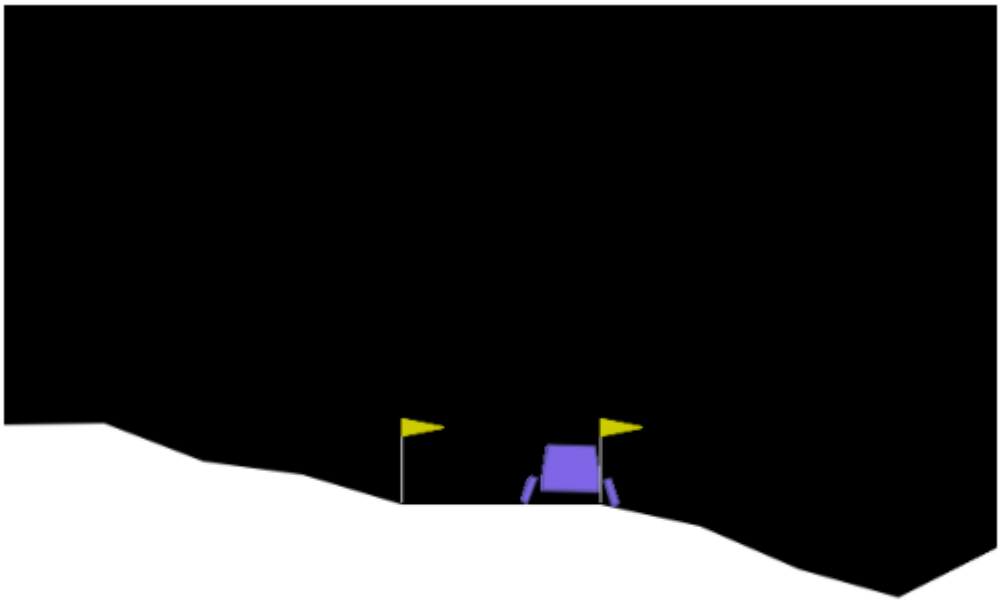


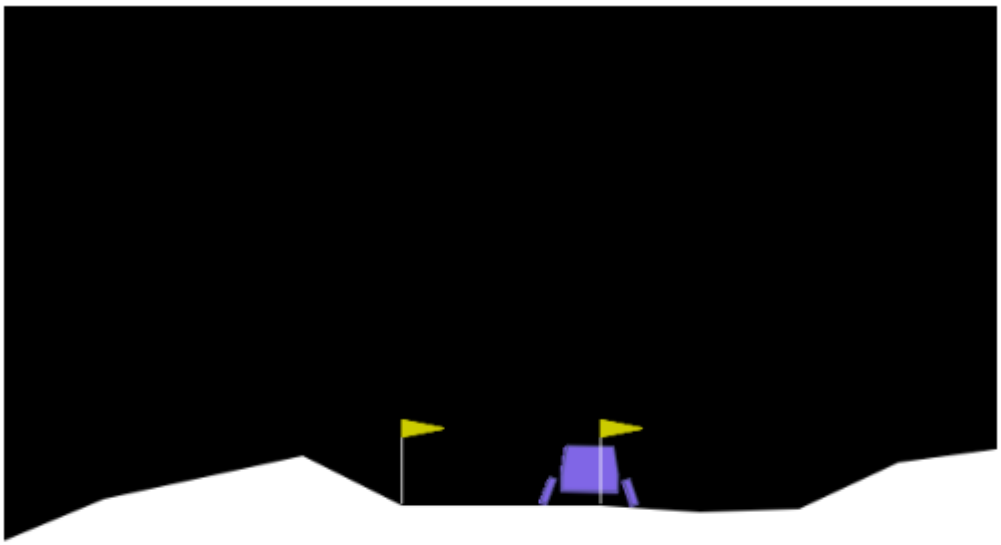


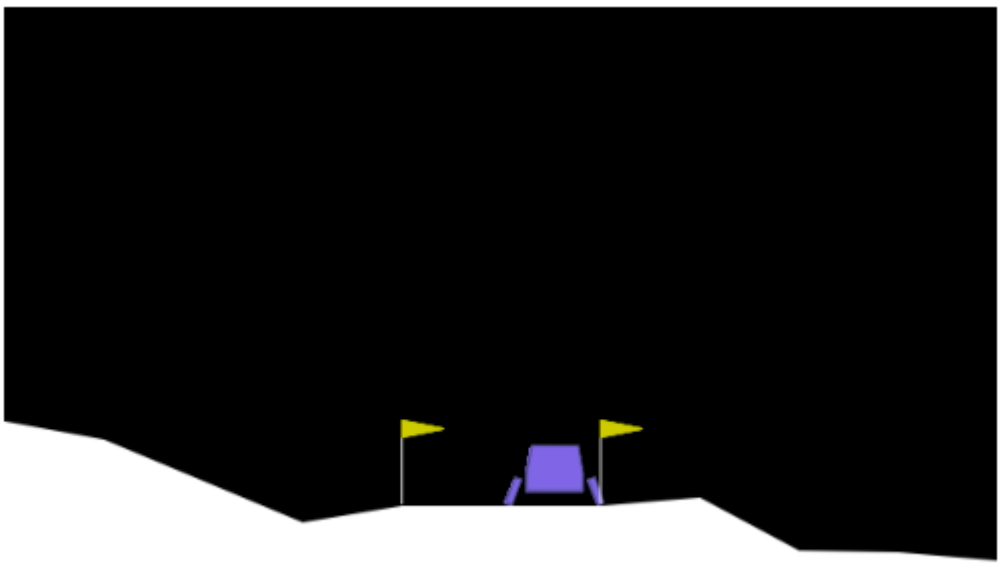


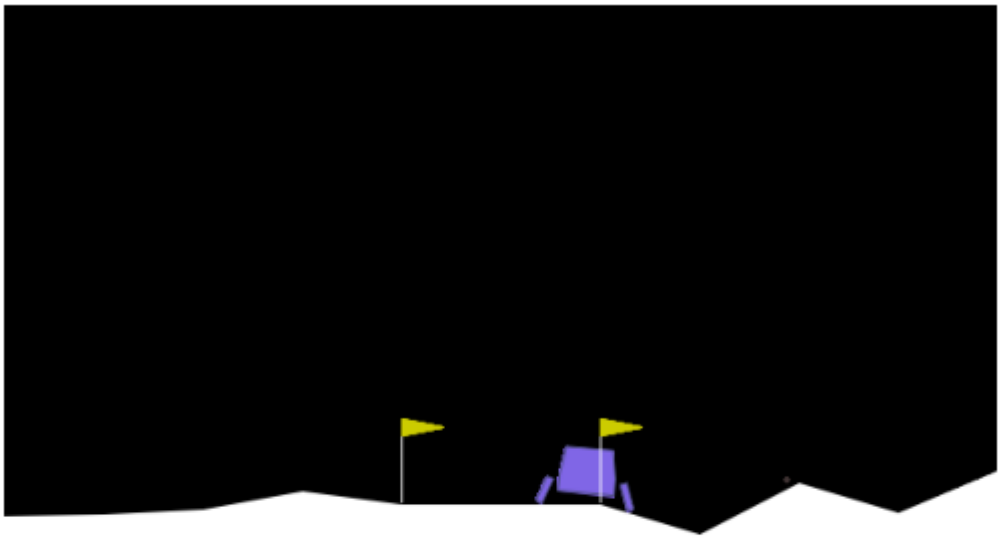




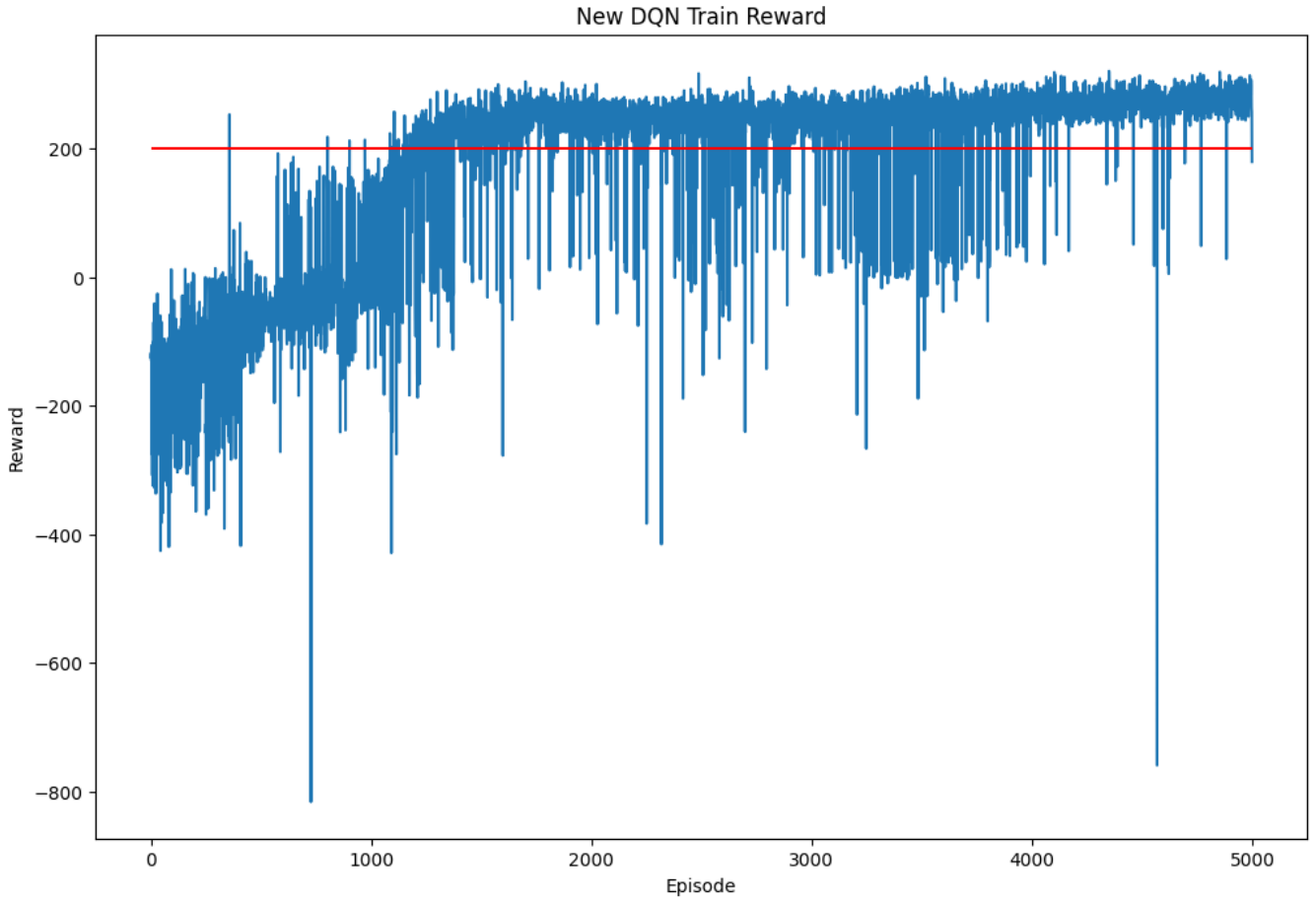




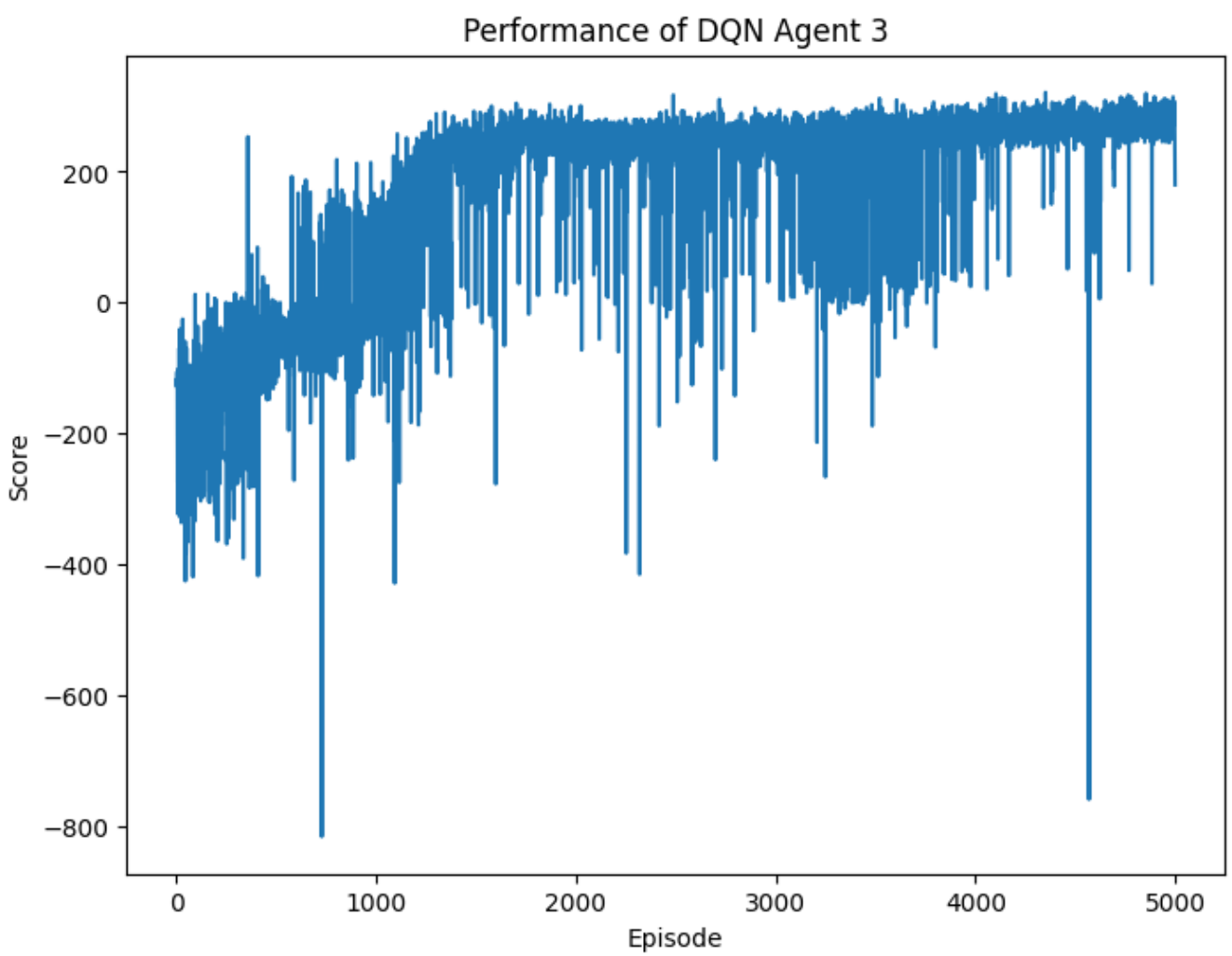




In [ ]:
```
plt.figure(figsize=(12,8))
plt.plot(scores, label='Reward')
plt.xlabel('Episode')
plt.ylabel('Reward')
plt.hlines(200, 0, len(scores), color='r')
plt.title("New DQN Train Reward")
plt.show()
```



In [ ]:
```
fig = plt.figure(figsize=(8,6))
ax = fig.add_subplot(111)
plt.plot(np.arange(len(scores)), scores)
plt.ylabel('Score')
plt.xlabel('Episode')
plt.title('Performance of DQN Agent 3')
plt.show()
```



In [ ]:
```
max_episodes = 1
master_frames, scores = train_model(max_episodes)
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
Cell In[14], line 2
      1 max_episodes = 1
----> 2 master_frames, scores = train_model(max_episodes)

NameError: name 'train_model' is not defined
```

In [ ]:
```
create_animation(master_frames[np.argmax(scores)])
```

Advantage Actor-critic

A2C, or Advantage Actor-Critic, is a reinforcement learning algorithm that merges the actor-critic and value-based methods. It is made up of two parts: an actor responsible for choosing actions and a critic responsible for evaluating the chosen actions. The actor is adjusted based on the advantage, which is the difference between the expected value of a specific state-action combination and the value of the state. The critic is updated based on the temporal difference error, which calculates the discrepancy between the estimated state value and the observed reward. A2C combines policy gradient methods to update the actor and value-based methods to update the critic, allowing it to balance exploration and exploitation, as well as accurately estimate values. A2C can be parallelized for efficient use in large environments.

Create environment for the A2C model

In [ ]:
```
train_env = gym.make('LunarLander-v2',
                     continuous=False,
                     gravity=-10.0,
```

```
                enable_wind=False,
                wind_power=15.0,
                turbulence_power=1.5
                )
train_env.seed(0)

test_env = gym.make('LunarLander-v2',
                continuous=False,
                gravity=-10.0,
                enable_wind=False,
                wind_power=15.0,
                turbulence_power=1.5
                )
test_env.seed(1)
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:317: DeprecationWarni
ng: WARN: Initializing wrapper in old step API which returns one bool instead of two. It is r
ecommended to set `new_step_api=True` to use new step API. This will be the default behaviour
in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\wrappers\step_api_compatibili
ty.py:39: DeprecationWarning: WARN: Initializing environment in old step API which returns on
e bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This
will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:256: DeprecationWarni
ng: WARN: Function `env.seed(seed)` is marked as deprecated and will be removed in the futur
e. Please use `env.reset(seed=seed)` instead.
  deprecation(
```

Out[ ]:
```
[1]
```

Setup model architecture

In [ ]:
```
class MLP(nn.Module):
    def __init__(self, input_dim, hidden_dim, output_dim, dropout = 0.1):
        super().__init__()

        self.net = nn.Sequential(
            nn.Linear(input_dim, hidden_dim),
            nn.Dropout(dropout),
            nn.PReLU(),
            nn.Linear(hidden_dim, hidden_dim),
            nn.Dropout(dropout),
            nn.PReLU(),
            nn.Linear(hidden_dim, output_dim)
        )

    def forward(self, x):
        x = self.net(x)
        return x
```

In [ ]:
```
class ActorCritic(nn.Module):
    def __init__(self, actor, critic):
        super().__init__()

        self.actor = actor
        self.critic = critic

    def forward(self, state):

        action_pred = self.actor(state)
        value_pred = self.critic(state)

        return action_pred, value_pred
```

In [ ]:
```
INPUT_DIM = train_env.observation_space.shape[0]
HIDDEN_DIM = 128
```

```
OUTPUT_DIM = test_env.action_space.n

actor = MLP(INPUT_DIM, HIDDEN_DIM, OUTPUT_DIM)
critic = MLP(INPUT_DIM, HIDDEN_DIM, 1)

a2c_policy = ActorCritic(actor, critic)
```

In [ ]:
```
def init_weights(m):
    if type(m) == nn.Linear:
        torch.nn.init.xavier_normal_(m.weight)
        m.bias.data.fill_(0)
```

In [ ]:
```
a2c_policy.apply(init_weights)
```

Out[ ]:
```
ActorCritic(
  (actor): MLP(
    (net): Sequential(
      (0): Linear(in_features=8, out_features=128, bias=True)
      (1): Dropout(p=0.1, inplace=False)
      (2): PReLU(num_parameters=1)
      (3): Linear(in_features=128, out_features=128, bias=True)
      (4): Dropout(p=0.1, inplace=False)
      (5): PReLU(num_parameters=1)
      (6): Linear(in_features=128, out_features=4, bias=True)
    )
  )
  (critic): MLP(
    (net): Sequential(
      (0): Linear(in_features=8, out_features=128, bias=True)
      (1): Dropout(p=0.1, inplace=False)
      (2): PReLU(num_parameters=1)
      (3): Linear(in_features=128, out_features=128, bias=True)
      (4): Dropout(p=0.1, inplace=False)
      (5): PReLU(num_parameters=1)
      (6): Linear(in_features=128, out_features=1, bias=True)
    )
  )
)
```

In [ ]:
```
LEARNING_RATE = 0.0005

optimizer = optim.Adam(a2c_policy.parameters(), lr=LEARNING_RATE)
```

In [ ]:
```
def calculate_returns(rewards, discount_factor, normalize=True):
    returns = []
    R = 0
    for r in reversed(rewards):
        R = r + R * discount_factor
        returns.insert(0, R)
    returns = torch.tensor(returns)
    if normalize:
        returns = (returns - returns.mean()) / returns.std()
    return returns
```

In [ ]:
```
def calculate_advantages(returns, values, normalize=True):
    advantages = returns - values
    if normalize:
        advantages = (advantages - advantages.mean()) / advantages.std()
    return advantages
```

In [ ]:
```
def update_policy(advantages, log_prob_actions, returns, values, optimizer):
    advantages = advantages.detach()
    returns = returns.detach()
    policy_loss = - (advantages * log_prob_actions).sum()
    value_loss = F.smooth_l1_loss(returns, values).sum()
    optimizer.zero_grad()
    policy_loss.backward()
```

```
    value_loss.backward()
    optimizer.step()
    return policy_loss.item(), value_loss.item()
```

Training A2C Model

In [ ]:
```
def a2c_train(env, policy, optimizer, discount_factor):
    policy.train()
    log_prob_actions = []
    values = []
    rewards = []
    done = False
    episode_reward = 0
    state = env.reset()
    while not done:
        state = torch.FloatTensor(state).unsqueeze(0)
        action_pred = actor(state)
        value_pred = critic(state)
        action_prob = F.softmax(action_pred, dim = -1)
        dist = distributions.Categorical(action_prob)
        action = dist.sample()
        log_prob_action = dist.log_prob(action)
        state, reward, done, _ = env.step(action.item())
        log_prob_actions.append(log_prob_action)
        values.append(value_pred)
        rewards.append(reward)
        episode_reward += reward
    log_prob_actions = torch.cat(log_prob_actions)
    values = torch.cat(values).squeeze(-1)
    returns = calculate_returns(rewards, discount_factor)
    advantages = calculate_advantages(returns, values)
    policy_loss, value_loss = update_policy(advantages, log_prob_actions, returns, values, op
    return policy_loss, value_loss, episode_reward
```

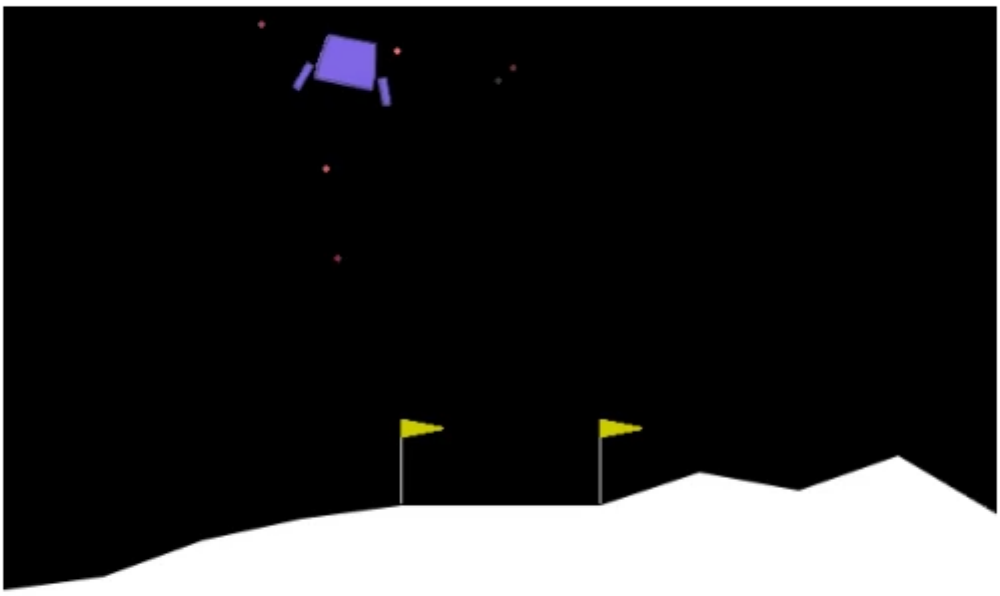
In [ ]:
```
def a2c_evaluate(env, policy):
    policy.eval()
    rewards = []
    done = False
    episode_reward = 0
    frames = []
    state = env.reset()
    while not done:
        state = torch.FloatTensor(state).unsqueeze(0)
        with torch.no_grad():
            action_pred, _ = policy(state)
            action_prob = F.softmax(action_pred, dim=-1)
        action = torch.argmax(action_prob, dim=-1)
        state, reward, done, _ = env.step(action.item())
        screen = env.render(mode='rgb_array')
        episode_reward += reward
        frames.append(screen)
    return episode_reward, frames
```

In [ ]:
```
MAX_EPISODES = 5_000
DISCOUNT_FACTOR = 0.99
N_TRIALS = 25
REWARD_THRESHOLD = 200
PRINT_EVERY = 10
VIDEO_EVERY = 750
a2c_train_rewards = []
a2c_test_rewards = []
for episode in range(1, MAX_EPISODES+1):
    a2c_policy_loss, a2c_value_loss, a2c_train_reward = a2c_train(
        train_env, a2c_policy, optimizer, DISCOUNT_FACTOR)
    a2c_test_reward, frames = a2c_evaluate(test_env, a2c_policy)
    if episode % VIDEO_EVERY == 0 or a2c_test_reward >= REWARD_THRESHOLD + 100:
        create_animation(frames, f"./videos/A2C/A2C-{episode}-{a2c_test_reward}.mp4")
```

```
    else:
        del frames
    a2c_train_rewards.append(a2c_train_reward)
    a2c_test_rewards.append(a2c_test_reward)
    mean_a2c_train_rewards = np.mean(a2c_train_rewards[-N_TRIALS:])
    mean_a2c_test_rewards = np.mean(a2c_test_rewards[-N_TRIALS:])
    if episode % PRINT_EVERY == 0:
        print(
            f'| Episode: {episode:3} | Mean Train Rewards: {mean_a2c_train_rewards:7.1f} | Me
    if a2c_train_reward >= REWARD_THRESHOLD:
        print(f'Reached reward train threshold in {episode} episodes')
        torch.save(a2c_policy.state_dict(),
                   f"./checkpoints/A2C/A2CTrain-{episode}.pth")
    if a2c_test_reward >= REWARD_THRESHOLD:
        print(f'Reached reward test threshold in {episode} episodes')
        torch.save(a2c_policy.state_dict(),
                   f"./checkpoints/A2C/A2CTest-{episode}.pth")
```

```
| Episode:  10 | Mean Train Rewards:  -193.9 | Mean Test Rewards:  -178.5 |
| Episode:  20 | Mean Train Rewards:  -179.5 | Mean Test Rewards:  -395.7 |
| Episode:  30 | Mean Train Rewards:  -189.9 | Mean Test Rewards:  -460.9 |
| Episode:  40 | Mean Train Rewards:  -190.1 | Mean Test Rewards:  -421.6 |
| Episode:  50 | Mean Train Rewards:  -163.8 | Mean Test Rewards:  -345.0 |
| Episode:  60 | Mean Train Rewards:  -157.2 | Mean Test Rewards:  -490.5 |
| Episode:  70 | Mean Train Rewards:  -159.9 | Mean Test Rewards:  -555.3 |
| Episode:  80 | Mean Train Rewards:  -157.0 | Mean Test Rewards:  -674.7 |
| Episode:  90 | Mean Train Rewards:  -151.7 | Mean Test Rewards:  -766.8 |
| Episode: 100 | Mean Train Rewards:  -152.3 | Mean Test Rewards:  -816.9 |
| Episode: 110 | Mean Train Rewards:  -147.1 | Mean Test Rewards:  -572.3 |
| Episode: 120 | Mean Train Rewards:  -129.6 | Mean Test Rewards:  -635.6 |
| Episode: 130 | Mean Train Rewards:  -144.0 | Mean Test Rewards:  -789.7 |
| Episode: 140 | Mean Train Rewards:  -164.8 | Mean Test Rewards:  -943.8 |
| Episode: 150 | Mean Train Rewards:  -170.8 | Mean Test Rewards: -1100.8 |
| Episode: 160 | Mean Train Rewards:  -164.1 | Mean Test Rewards: -1026.0 |
| Episode: 170 | Mean Train Rewards:  -166.5 | Mean Test Rewards: -1041.6 |
| Episode: 180 | Mean Train Rewards:  -159.2 | Mean Test Rewards: -1075.1 |
| Episode: 190 | Mean Train Rewards:  -148.0 | Mean Test Rewards: -1337.3 |
| Episode: 200 | Mean Train Rewards:  -136.3 | Mean Test Rewards: -1344.5 |
| Episode: 210 | Mean Train Rewards:  -160.3 | Mean Test Rewards: -1418.5 |
| Episode: 220 | Mean Train Rewards:  -173.0 | Mean Test Rewards: -1389.3 |
| Episode: 230 | Mean Train Rewards:  -177.7 | Mean Test Rewards: -1345.8 |
| Episode: 240 | Mean Train Rewards:  -166.8 | Mean Test Rewards: -1174.2 |
| Episode: 250 | Mean Train Rewards:  -155.4 | Mean Test Rewards: -1008.0 |
| Episode: 260 | Mean Train Rewards:  -177.8 | Mean Test Rewards: -1018.3 |
| Episode: 270 | Mean Train Rewards:  -129.5 | Mean Test Rewards: -1199.3 |
| Episode: 280 | Mean Train Rewards:   -86.8 | Mean Test Rewards: -1333.5 |
| Episode: 290 | Mean Train Rewards:   -91.2 | Mean Test Rewards: -1450.5 |
| Episode: 300 | Mean Train Rewards:   -95.2 | Mean Test Rewards: -1798.9 |
| Episode: 310 | Mean Train Rewards:   -93.6 | Mean Test Rewards: -1514.9 |
| Episode: 320 | Mean Train Rewards:   -63.9 | Mean Test Rewards: -1403.5 |
| Episode: 330 | Mean Train Rewards:   -59.6 | Mean Test Rewards: -1076.5 |
| Episode: 340 | Mean Train Rewards:   -51.8 | Mean Test Rewards:  -934.4 |
| Episode: 350 | Mean Train Rewards:   -24.2 | Mean Test Rewards:  -811.1 |
| Episode: 360 | Mean Train Rewards:   -38.2 | Mean Test Rewards:  -793.6 |
| Episode: 370 | Mean Train Rewards:   -40.0 | Mean Test Rewards:  -692.0 |
| Episode: 380 | Mean Train Rewards:   -46.7 | Mean Test Rewards:  -593.2 |
| Episode: 390 | Mean Train Rewards:   -45.3 | Mean Test Rewards:  -499.7 |
| Episode: 400 | Mean Train Rewards:   -75.0 | Mean Test Rewards:  -578.0 |
| Episode: 410 | Mean Train Rewards:  -111.8 | Mean Test Rewards:  -600.8 |
| Episode: 420 | Mean Train Rewards:  -138.9 | Mean Test Rewards:  -833.3 |
| Episode: 430 | Mean Train Rewards:  -131.3 | Mean Test Rewards:  -842.9 |
| Episode: 440 | Mean Train Rewards:   -72.7 | Mean Test Rewards:  -828.9 |
| Episode: 450 | Mean Train Rewards:   -14.9 | Mean Test Rewards:  -552.7 |
| Episode: 460 | Mean Train Rewards:     8.5 | Mean Test Rewards:  -349.7 |
| Episode: 470 | Mean Train Rewards:    13.3 | Mean Test Rewards:  -188.3 |
| Episode: 480 | Mean Train Rewards:    15.6 | Mean Test Rewards:  -113.6 |
| Episode: 490 | Mean Train Rewards:    24.4 | Mean Test Rewards:  -117.9 |
| Episode: 500 | Mean Train Rewards:    15.4 | Mean Test Rewards:  -127.2 |
| Episode: 510 | Mean Train Rewards:     1.2 | Mean Test Rewards:  -136.8 |
| Episode: 520 | Mean Train Rewards:     8.4 | Mean Test Rewards:  -140.8 |
| Episode: 530 | Mean Train Rewards:    24.1 | Mean Test Rewards:  -176.3 |
| Episode: 540 | Mean Train Rewards:    20.2 | Mean Test Rewards:  -215.9 |
| Episode: 550 | Mean Train Rewards:   -15.6 | Mean Test Rewards:  -201.2 |
| Episode: 560 | Mean Train Rewards:    29.9 | Mean Test Rewards:  -162.7 |
| Episode: 570 | Mean Train Rewards:    43.2 | Mean Test Rewards:  -125.0 |
Reached reward train threshold in 571 episodes
| Episode: 580 | Mean Train Rewards:    38.9 | Mean Test Rewards:  -128.0 |
Reached reward train threshold in 585 episodes
| Episode: 590 | Mean Train Rewards:    41.8 | Mean Test Rewards:  -129.4 |
| Episode: 600 | Mean Train Rewards:    16.6 | Mean Test Rewards:  -125.9 |
Reached reward train threshold in 604 episodes
Reached reward train threshold in 609 episodes
| Episode: 610 | Mean Train Rewards:    17.1 | Mean Test Rewards:  -116.8 |
Reached reward train threshold in 619 episodes
| Episode: 620 | Mean Train Rewards:    43.3 | Mean Test Rewards:  -101.6 |
Reached reward train threshold in 625 episodes
| Episode: 630 | Mean Train Rewards:    58.3 | Mean Test Rewards:   -80.3 |
```

```
Reached reward train threshold in 634 episodes
Reached reward train threshold in 635 episodes
| Episode: 640 | Mean Train Rewards:    66.8 | Mean Test Rewards:   -88.7 |
Reached reward train threshold in 640 episodes
Reached reward train threshold in 641 episodes
Reached reward train threshold in 649 episodes
| Episode: 650 | Mean Train Rewards:    57.9 | Mean Test Rewards:   -84.3 |
Reached reward test threshold in 652 episodes
Reached reward train threshold in 655 episodes
Reached reward train threshold in 658 episodes
| Episode: 660 | Mean Train Rewards:    37.1 | Mean Test Rewards:   -63.1 |
Reached reward train threshold in 668 episodes
| Episode: 670 | Mean Train Rewards:    31.1 | Mean Test Rewards:   -44.0 |
Reached reward train threshold in 678 episodes
| Episode: 680 | Mean Train Rewards:    47.3 | Mean Test Rewards:   -49.8 |
Reached reward train threshold in 682 episodes
Reached reward train threshold in 683 episodes
Reached reward train threshold in 685 episodes
Reached reward train threshold in 686 episodes
| Episode: 690 | Mean Train Rewards:    64.2 | Mean Test Rewards:   -53.1 |
Reached reward train threshold in 697 episodes
| Episode: 700 | Mean Train Rewards:    65.8 | Mean Test Rewards:   -40.1 |
Reached reward train threshold in 703 episodes
Reached reward train threshold in 708 episodes
| Episode: 710 | Mean Train Rewards:    72.2 | Mean Test Rewards:   -45.0 |
Reached reward train threshold in 711 episodes
Reached reward train threshold in 712 episodes
Reached reward train threshold in 716 episodes
| Episode: 720 | Mean Train Rewards:    67.0 | Mean Test Rewards:   -66.0 |
Reached reward train threshold in 722 episodes
Reached reward train threshold in 723 episodes
Reached reward train threshold in 728 episodes
| Episode: 730 | Mean Train Rewards:    40.2 | Mean Test Rewards:   -75.2 |
Reached reward train threshold in 734 episodes
Reached reward train threshold in 737 episodes
| Episode: 740 | Mean Train Rewards:    20.3 | Mean Test Rewards:  -105.5 |
```



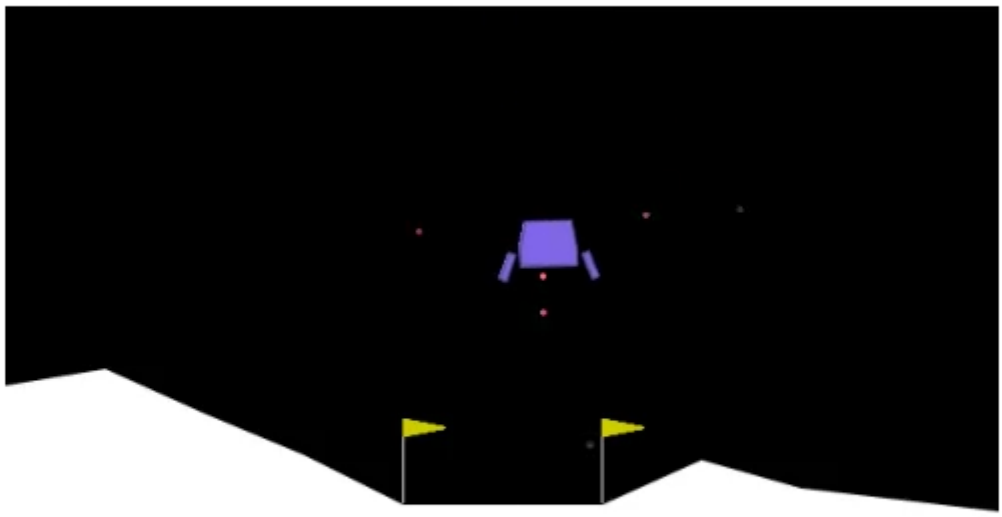
```
| Episode: 750 | Mean Train Rewards:     6.5 | Mean Test Rewards:  -138.3 |
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

```
Reached reward train threshold in 759 episodes
| Episode: 760 | Mean Train Rewards:    22.2 | Mean Test Rewards:  -156.2 |
Reached reward train threshold in 765 episodes
| Episode: 770 | Mean Train Rewards:    19.3 | Mean Test Rewards:  -181.4 |
Reached reward train threshold in 778 episodes
| Episode: 780 | Mean Train Rewards:    21.8 | Mean Test Rewards:  -172.6 |
Reached reward train threshold in 786 episodes
Reached reward train threshold in 789 episodes
| Episode: 790 | Mean Train Rewards:    54.4 | Mean Test Rewards:  -150.9 |
Reached reward train threshold in 791 episodes
Reached reward train threshold in 794 episodes
Reached reward train threshold in 799 episodes
| Episode: 800 | Mean Train Rewards:   103.9 | Mean Test Rewards:  -118.1 |
Reached reward train threshold in 800 episodes
Reached reward train threshold in 805 episodes
| Episode: 810 | Mean Train Rewards:    92.8 | Mean Test Rewards:  -107.8 |
Reached reward train threshold in 812 episodes
Reached reward train threshold in 816 episodes
| Episode: 820 | Mean Train Rewards:    88.4 | Mean Test Rewards:  -111.2 |
| Episode: 830 | Mean Train Rewards:    72.8 | Mean Test Rewards:  -142.8 |
| Episode: 840 | Mean Train Rewards:    91.2 | Mean Test Rewards:  -139.0 |
Reached reward train threshold in 840 episodes
Reached reward train threshold in 844 episodes
| Episode: 850 | Mean Train Rewards:    86.6 | Mean Test Rewards:  -109.1 |
Reached reward train threshold in 859 episodes
| Episode: 860 | Mean Train Rewards:    68.4 | Mean Test Rewards:   -43.4 |
Reached reward test threshold in 861 episodes
Reached reward train threshold in 862 episodes
| Episode: 870 | Mean Train Rewards:    57.5 | Mean Test Rewards:    -4.9 |
Reached reward train threshold in 871 episodes
| Episode: 880 | Mean Train Rewards:    54.6 | Mean Test Rewards:    -3.1 |
Reached reward train threshold in 885 episodes
| Episode: 890 | Mean Train Rewards:    64.5 | Mean Test Rewards:   -31.6 |
Reached reward train threshold in 891 episodes
Reached reward train threshold in 893 episodes
Reached reward train threshold in 896 episodes
Reached reward test threshold in 899 episodes
| Episode: 900 | Mean Train Rewards:    99.2 | Mean Test Rewards:   -29.8 |
Reached reward train threshold in 906 episodes
Reached reward train threshold in 908 episodes
Reached reward train threshold in 909 episodes
| Episode: 910 | Mean Train Rewards:   114.8 | Mean Test Rewards:   -22.4 |
Reached reward train threshold in 912 episodes
Reached reward train threshold in 913 episodes
Reached reward train threshold in 918 episodes
| Episode: 920 | Mean Train Rewards:   109.9 | Mean Test Rewards:   -10.7 |
Reached reward train threshold in 920 episodes
Reached reward train threshold in 921 episodes
Reached reward train threshold in 929 episodes
| Episode: 930 | Mean Train Rewards:   119.0 | Mean Test Rewards:   -11.9 |
Reached reward train threshold in 932 episodes
Reached reward train threshold in 935 episodes
| Episode: 940 | Mean Train Rewards:   106.7 | Mean Test Rewards:   -17.3 |
| Episode: 950 | Mean Train Rewards:    96.2 | Mean Test Rewards:   -25.5 |
| Episode: 960 | Mean Train Rewards:    95.3 | Mean Test Rewards:   -28.2 |
| Episode: 970 | Mean Train Rewards:   106.5 | Mean Test Rewards:   -24.4 |
Reached reward train threshold in 970 episodes
| Episode: 980 | Mean Train Rewards:   109.9 | Mean Test Rewards:   -27.0 |
| Episode: 990 | Mean Train Rewards:    88.4 | Mean Test Rewards:   -22.3 |
Reached reward train threshold in 999 episodes
| Episode: 1000 | Mean Train Rewards:    71.2 | Mean Test Rewards:    -1.4 |
Reached reward test threshold in 1000 episodes
| Episode: 1010 | Mean Train Rewards:    73.0 | Mean Test Rewards:     0.8 |
Reached reward train threshold in 1017 episodes
| Episode: 1020 | Mean Train Rewards:    94.6 | Mean Test Rewards:     4.3 |
| Episode: 1030 | Mean Train Rewards:   106.3 | Mean Test Rewards:    -5.3 |
Reached reward train threshold in 1033 episodes
| Episode: 1040 | Mean Train Rewards:   126.7 | Mean Test Rewards:   -16.6 |
Reached reward train threshold in 1044 episodes
```

```
| Episode: 1050 | Mean Train Rewards:   130.3 | Mean Test Rewards:   -25.5 |
| Episode: 1060 | Mean Train Rewards:   130.9 | Mean Test Rewards:   -33.3 |
| Episode: 1070 | Mean Train Rewards:   104.2 | Mean Test Rewards:   -29.9 |
Reached reward train threshold in 1075 episodes
| Episode: 1080 | Mean Train Rewards:    95.0 | Mean Test Rewards:   -23.1 |
| Episode: 1090 | Mean Train Rewards:   110.9 | Mean Test Rewards:   -15.2 |
| Episode: 1100 | Mean Train Rewards:   104.5 | Mean Test Rewards:   -12.7 |
Reached reward train threshold in 1106 episodes
| Episode: 1110 | Mean Train Rewards:   125.5 | Mean Test Rewards:   -14.2 |
Reached reward train threshold in 1110 episodes
Reached reward train threshold in 1118 episodes
| Episode: 1120 | Mean Train Rewards:   136.4 | Mean Test Rewards:   -17.5 |
Reached reward train threshold in 1129 episodes
| Episode: 1130 | Mean Train Rewards:   132.0 | Mean Test Rewards:   -14.6 |
| Episode: 1140 | Mean Train Rewards:   127.5 | Mean Test Rewards:   -10.7 |
Reached reward train threshold in 1147 episodes
| Episode: 1150 | Mean Train Rewards:   125.8 | Mean Test Rewards:     3.1 |
Reached reward train threshold in 1150 episodes
Reached reward test threshold in 1153 episodes
Reached reward test threshold in 1154 episodes
Reached reward test threshold in 1156 episodes
Reached reward test threshold in 1158 episodes
Reached reward test threshold in 1159 episodes
| Episode: 1160 | Mean Train Rewards:   113.4 | Mean Test Rewards:    83.2 |
Reached reward train threshold in 1160 episodes
Reached reward test threshold in 1160 episodes
Reached reward train threshold in 1161 episodes
Reached reward train threshold in 1163 episodes
Reached reward train threshold in 1167 episodes
Reached reward test threshold in 1167 episodes
Reached reward test threshold in 1169 episodes
| Episode: 1170 | Mean Train Rewards:   126.3 | Mean Test Rewards:   137.7 |
Reached reward train threshold in 1170 episodes
Reached reward train threshold in 1171 episodes
Reached reward train threshold in 1174 episodes
Reached reward test threshold in 1175 episodes
Reached reward train threshold in 1176 episodes
| Episode: 1180 | Mean Train Rewards:   128.4 | Mean Test Rewards:   148.8 |
Reached reward test threshold in 1180 episodes
Reached reward test threshold in 1181 episodes
Reached reward test threshold in 1183 episodes
Reached reward train threshold in 1185 episodes
Reached reward test threshold in 1186 episodes
| Episode: 1190 | Mean Train Rewards:   120.7 | Mean Test Rewards:   142.8 |
Reached reward test threshold in 1190 episodes
Reached reward test threshold in 1191 episodes
Reached reward test threshold in 1192 episodes
Reached reward test threshold in 1195 episodes
Reached reward test threshold in 1196 episodes
Reached reward test threshold in 1198 episodes
| Episode: 1200 | Mean Train Rewards:    93.7 | Mean Test Rewards:   148.0 |
| Episode: 1210 | Mean Train Rewards:    89.5 | Mean Test Rewards:   113.2 |
Reached reward train threshold in 1213 episodes
Reached reward train threshold in 1219 episodes
| Episode: 1220 | Mean Train Rewards:   127.0 | Mean Test Rewards:    58.2 |
Reached reward test threshold in 1222 episodes
Reached reward train threshold in 1225 episodes
Reached reward train threshold in 1226 episodes
Reached reward test threshold in 1228 episodes
| Episode: 1230 | Mean Train Rewards:   130.0 | Mean Test Rewards:    65.6 |
Reached reward test threshold in 1231 episodes
Reached reward train threshold in 1232 episodes
Reached reward test threshold in 1233 episodes
| Episode: 1240 | Mean Train Rewards:   130.1 | Mean Test Rewards:    89.1 |
Reached reward train threshold in 1243 episodes
Reached reward train threshold in 1244 episodes
| Episode: 1250 | Mean Train Rewards:   127.0 | Mean Test Rewards:    59.4 |
Reached reward train threshold in 1254 episodes
| Episode: 1260 | Mean Train Rewards:   110.7 | Mean Test Rewards:     5.7 |
```

```
Reached reward train threshold in 1265 episodes
Reached reward train threshold in 1268 episodes
| Episode: 1270 | Mean Train Rewards:   116.2 | Mean Test Rewards:    -5.4 |
| Episode: 1280 | Mean Train Rewards:   111.0 | Mean Test Rewards:    -8.7 |
Reached reward train threshold in 1286 episodes
| Episode: 1290 | Mean Train Rewards:   115.1 | Mean Test Rewards:   -18.5 |
| Episode: 1300 | Mean Train Rewards:   132.0 | Mean Test Rewards:    -8.6 |
Reached reward test threshold in 1300 episodes
| Episode: 1310 | Mean Train Rewards:   140.5 | Mean Test Rewards:     1.0 |
Reached reward train threshold in 1313 episodes
| Episode: 1320 | Mean Train Rewards:   144.7 | Mean Test Rewards:    -1.9 |
| Episode: 1330 | Mean Train Rewards:   144.0 | Mean Test Rewards:    -9.8 |
| Episode: 1340 | Mean Train Rewards:   137.4 | Mean Test Rewards:     6.2 |
Reached reward train threshold in 1346 episodes
Reached reward test threshold in 1347 episodes
| Episode: 1350 | Mean Train Rewards:   142.9 | Mean Test Rewards:    50.1 |
| Episode: 1360 | Mean Train Rewards:   142.4 | Mean Test Rewards:    68.8 |
| Episode: 1370 | Mean Train Rewards:   128.6 | Mean Test Rewards:    98.6 |
Reached reward train threshold in 1371 episodes
Reached reward test threshold in 1372 episodes
Reached reward test threshold in 1373 episodes
| Episode: 1380 | Mean Train Rewards:   130.8 | Mean Test Rewards:   127.1 |
Reached reward train threshold in 1380 episodes
Reached reward test threshold in 1380 episodes
Reached reward test threshold in 1383 episodes
| Episode: 1390 | Mean Train Rewards:   125.2 | Mean Test Rewards:   132.4 |
| Episode: 1400 | Mean Train Rewards:   122.2 | Mean Test Rewards:   123.8 |
| Episode: 1410 | Mean Train Rewards:   116.7 | Mean Test Rewards:   102.3 |
Reached reward train threshold in 1410 episodes
| Episode: 1420 | Mean Train Rewards:   120.7 | Mean Test Rewards:    84.9 |
| Episode: 1430 | Mean Train Rewards:   123.9 | Mean Test Rewards:    70.9 |
| Episode: 1440 | Mean Train Rewards:   131.9 | Mean Test Rewards:    86.0 |
| Episode: 1450 | Mean Train Rewards:   134.6 | Mean Test Rewards:   100.9 |
Reached reward test threshold in 1451 episodes
Reached reward test threshold in 1456 episodes
| Episode: 1460 | Mean Train Rewards:   128.0 | Mean Test Rewards:   125.4 |
Reached reward test threshold in 1464 episodes
| Episode: 1470 | Mean Train Rewards:   124.9 | Mean Test Rewards:   122.5 |
Reached reward test threshold in 1473 episodes
Reached reward train threshold in 1474 episodes
Reached reward test threshold in 1475 episodes
Reached reward test threshold in 1476 episodes
Reached reward train threshold in 1478 episodes
| Episode: 1480 | Mean Train Rewards:   132.7 | Mean Test Rewards:   135.0 |
Reached reward test threshold in 1480 episodes
Reached reward test threshold in 1481 episodes
Reached reward test threshold in 1483 episodes
Reached reward test threshold in 1487 episodes
| Episode: 1490 | Mean Train Rewards:   129.3 | Mean Test Rewards:   136.6 |
Reached reward train threshold in 1491 episodes
Reached reward train threshold in 1496 episodes
Reached reward test threshold in 1496 episodes
Reached reward test threshold in 1498 episodes
Reached reward test threshold in 1499 episodes
```



```
| Episode: 1500 | Mean Train Rewards:   130.5 | Mean Test Rewards:   156.1 |
Reached reward test threshold in 1500 episodes
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

```
Reached reward test threshold in 1501 episodes
Reached reward test threshold in 1504 episodes
Reached reward train threshold in 1506 episodes
Reached reward test threshold in 1506 episodes
Reached reward test threshold in 1507 episodes
Reached reward test threshold in 1508 episodes
| Episode: 1510 | Mean Train Rewards:   113.9 | Mean Test Rewards:   165.0 |
Reached reward test threshold in 1512 episodes
Reached reward test threshold in 1516 episodes
| Episode: 1520 | Mean Train Rewards:   104.0 | Mean Test Rewards:   182.4 |
Reached reward test threshold in 1523 episodes
Reached reward test threshold in 1526 episodes
Reached reward test threshold in 1528 episodes
Reached reward test threshold in 1529 episodes
| Episode: 1530 | Mean Train Rewards:   106.6 | Mean Test Rewards:   177.9 |
Reached reward test threshold in 1531 episodes
| Episode: 1540 | Mean Train Rewards:   112.5 | Mean Test Rewards:   132.9 |
| Episode: 1550 | Mean Train Rewards:   126.6 | Mean Test Rewards:    64.5 |
| Episode: 1560 | Mean Train Rewards:   131.2 | Mean Test Rewards:    -2.1 |
| Episode: 1570 | Mean Train Rewards:   138.7 | Mean Test Rewards:     7.4 |
| Episode: 1580 | Mean Train Rewards:   139.1 | Mean Test Rewards:    51.8 |
Reached reward test threshold in 1581 episodes
Reached reward train threshold in 1583 episodes
Reached reward test threshold in 1589 episodes
| Episode: 1590 | Mean Train Rewards:   139.8 | Mean Test Rewards:   110.6 |
| Episode: 1600 | Mean Train Rewards:   129.3 | Mean Test Rewards:   130.2 |
Reached reward test threshold in 1605 episodes
Reached reward test threshold in 1606 episodes
Reached reward test threshold in 1608 episodes
| Episode: 1610 | Mean Train Rewards:   101.5 | Mean Test Rewards:   149.5 |
Reached reward test threshold in 1610 episodes
| Episode: 1620 | Mean Train Rewards:   100.3 | Mean Test Rewards:   140.2 |
Reached reward test threshold in 1629 episodes
| Episode: 1630 | Mean Train Rewards:    93.1 | Mean Test Rewards:   150.3 |
Reached reward test threshold in 1630 episodes
Reached reward test threshold in 1632 episodes
Reached reward test threshold in 1633 episodes
Reached reward train threshold in 1636 episodes
Reached reward test threshold in 1637 episodes
| Episode: 1640 | Mean Train Rewards:    80.2 | Mean Test Rewards:   164.8 |
Reached reward test threshold in 1640 episodes
Reached reward train threshold in 1643 episodes
Reached reward test threshold in 1647 episodes
Reached reward test threshold in 1648 episodes
| Episode: 1650 | Mean Train Rewards:    90.8 | Mean Test Rewards:   171.1 |
Reached reward test threshold in 1650 episodes
Reached reward test threshold in 1655 episodes
Reached reward test threshold in 1659 episodes
| Episode: 1660 | Mean Train Rewards:   106.9 | Mean Test Rewards:   150.7 |
Reached reward test threshold in 1661 episodes
Reached reward test threshold in 1662 episodes
Reached reward test threshold in 1666 episodes
Reached reward test threshold in 1668 episodes
| Episode: 1670 | Mean Train Rewards:   108.6 | Mean Test Rewards:   142.6 |
Reached reward train threshold in 1670 episodes
Reached reward test threshold in 1672 episodes
Reached reward train threshold in 1677 episodes
| Episode: 1680 | Mean Train Rewards:   110.5 | Mean Test Rewards:   126.1 |
Reached reward train threshold in 1682 episodes
| Episode: 1690 | Mean Train Rewards:   139.0 | Mean Test Rewards:    84.6 |
| Episode: 1700 | Mean Train Rewards:   143.2 | Mean Test Rewards:    52.0 |
Reached reward train threshold in 1703 episodes
| Episode: 1710 | Mean Train Rewards:   137.7 | Mean Test Rewards:    65.6 |
| Episode: 1720 | Mean Train Rewards:   136.7 | Mean Test Rewards:    97.2 |
Reached reward test threshold in 1724 episodes
Reached reward test threshold in 1729 episodes
| Episode: 1730 | Mean Train Rewards:   121.9 | Mean Test Rewards:   124.4 |
| Episode: 1740 | Mean Train Rewards:   117.5 | Mean Test Rewards:   131.8 |
Reached reward test threshold in 1741 episodes
```

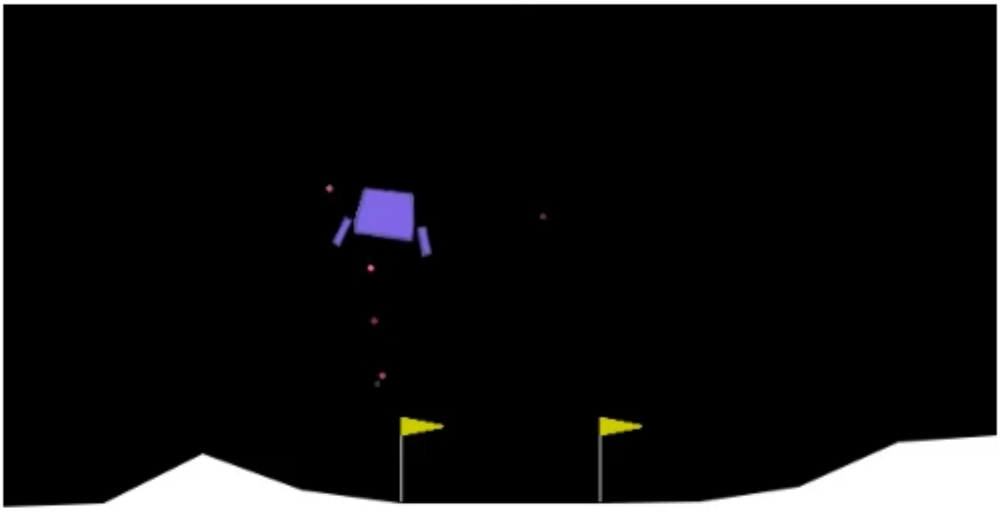
```
Reached reward test threshold in 1742 episodes
Reached reward test threshold in 1746 episodes
Reached reward test threshold in 1747 episodes
Reached reward test threshold in 1748 episodes
| Episode: 1750 | Mean Train Rewards:   101.9 | Mean Test Rewards:   162.2 |
Reached reward test threshold in 1750 episodes
Reached reward test threshold in 1751 episodes
Reached reward test threshold in 1753 episodes
Reached reward test threshold in 1758 episodes
| Episode: 1760 | Mean Train Rewards:    81.9 | Mean Test Rewards:   167.2 |
Reached reward train threshold in 1761 episodes
Reached reward test threshold in 1761 episodes
Reached reward test threshold in 1762 episodes
Reached reward test threshold in 1763 episodes
Reached reward test threshold in 1766 episodes
Reached reward test threshold in 1767 episodes
| Episode: 1770 | Mean Train Rewards:    76.6 | Mean Test Rewards:   169.9 |
Reached reward train threshold in 1771 episodes
Reached reward train threshold in 1772 episodes
| Episode: 1780 | Mean Train Rewards:   101.7 | Mean Test Rewards:   149.6 |
Reached reward train threshold in 1781 episodes
| Episode: 1790 | Mean Train Rewards:   119.6 | Mean Test Rewards:   131.9 |
Reached reward test threshold in 1792 episodes
Reached reward test threshold in 1793 episodes
Reached reward test threshold in 1794 episodes
Reached reward test threshold in 1795 episodes
Reached reward test threshold in 1796 episodes
Reached reward test threshold in 1797 episodes
Reached reward test threshold in 1798 episodes
Reached reward test threshold in 1799 episodes
| Episode: 1800 | Mean Train Rewards:   123.5 | Mean Test Rewards:   166.1 |
Reached reward test threshold in 1800 episodes
Reached reward train threshold in 1802 episodes
Reached reward train threshold in 1803 episodes
Reached reward test threshold in 1803 episodes
Reached reward test threshold in 1804 episodes
Reached reward test threshold in 1805 episodes
| Episode: 1810 | Mean Train Rewards:   128.5 | Mean Test Rewards:   179.0 |
Reached reward test threshold in 1812 episodes
| Episode: 1820 | Mean Train Rewards:   138.1 | Mean Test Rewards:   159.7 |
Reached reward test threshold in 1821 episodes
Reached reward train threshold in 1823 episodes
Reached reward test threshold in 1824 episodes
Reached reward test threshold in 1825 episodes
Reached reward test threshold in 1826 episodes
Reached reward test threshold in 1827 episodes
Reached reward test threshold in 1829 episodes
| Episode: 1830 | Mean Train Rewards:   141.2 | Mean Test Rewards:   151.5 |
Reached reward test threshold in 1831 episodes
Reached reward test threshold in 1835 episodes
Reached reward test threshold in 1836 episodes
Reached reward test threshold in 1839 episodes
| Episode: 1840 | Mean Train Rewards:   142.2 | Mean Test Rewards:   171.5 |
Reached reward test threshold in 1841 episodes
Reached reward test threshold in 1846 episodes
Reached reward test threshold in 1847 episodes
| Episode: 1850 | Mean Train Rewards:   134.6 | Mean Test Rewards:   175.5 |
Reached reward train threshold in 1856 episodes
| Episode: 1860 | Mean Train Rewards:   136.1 | Mean Test Rewards:   144.8 |
| Episode: 1870 | Mean Train Rewards:   136.6 | Mean Test Rewards:   136.9 |
| Episode: 1880 | Mean Train Rewards:   140.9 | Mean Test Rewards:   130.5 |
| Episode: 1890 | Mean Train Rewards:   143.5 | Mean Test Rewards:   130.3 |
Reached reward train threshold in 1891 episodes
Reached reward train threshold in 1895 episodes
Reached reward train threshold in 1898 episodes
| Episode: 1900 | Mean Train Rewards:   156.5 | Mean Test Rewards:   134.8 |
Reached reward train threshold in 1903 episodes
Reached reward test threshold in 1904 episodes
| Episode: 1910 | Mean Train Rewards:   141.7 | Mean Test Rewards:   146.6 |
```

```
Reached reward test threshold in 1910 episodes
Reached reward test threshold in 1911 episodes
Reached reward test threshold in 1913 episodes
Reached reward test threshold in 1916 episodes
Reached reward test threshold in 1917 episodes
Reached reward test threshold in 1918 episodes
Reached reward test threshold in 1919 episodes
| Episode: 1920 | Mean Train Rewards:   119.3 | Mean Test Rewards:   164.2 |
Reached reward test threshold in 1924 episodes
Reached reward train threshold in 1925 episodes
Reached reward test threshold in 1925 episodes
Reached reward train threshold in 1929 episodes
Reached reward test threshold in 1929 episodes
| Episode: 1930 | Mean Train Rewards:   124.4 | Mean Test Rewards:   165.7 |
Reached reward test threshold in 1930 episodes
Reached reward test threshold in 1932 episodes
Reached reward train threshold in 1933 episodes
Reached reward test threshold in 1933 episodes
Reached reward test threshold in 1936 episodes
Reached reward test threshold in 1937 episodes
Reached reward test threshold in 1938 episodes
Reached reward train threshold in 1939 episodes
Reached reward test threshold in 1939 episodes
| Episode: 1940 | Mean Train Rewards:   148.8 | Mean Test Rewards:   182.6 |
Reached reward train threshold in 1940 episodes
Reached reward test threshold in 1940 episodes
Reached reward test threshold in 1941 episodes
Reached reward test threshold in 1943 episodes
Reached reward test threshold in 1946 episodes
Reached reward test threshold in 1948 episodes
| Episode: 1950 | Mean Train Rewards:   134.0 | Mean Test Rewards:   164.3 |
Reached reward test threshold in 1951 episodes
Reached reward test threshold in 1952 episodes
Reached reward test threshold in 1953 episodes
Reached reward test threshold in 1954 episodes
Reached reward train threshold in 1959 episodes
| Episode: 1960 | Mean Train Rewards:   106.1 | Mean Test Rewards:   162.9 |
Reached reward test threshold in 1960 episodes
Reached reward test threshold in 1964 episodes
Reached reward train threshold in 1969 episodes
| Episode: 1970 | Mean Train Rewards:    98.5 | Mean Test Rewards:   157.6 |
Reached reward train threshold in 1977 episodes
| Episode: 1980 | Mean Train Rewards:   118.8 | Mean Test Rewards:   147.1 |
Reached reward test threshold in 1980 episodes
Reached reward train threshold in 1981 episodes
Reached reward test threshold in 1983 episodes
Reached reward train threshold in 1984 episodes
Reached reward test threshold in 1985 episodes
| Episode: 1990 | Mean Train Rewards:   108.7 | Mean Test Rewards:   153.0 |
| Episode: 2000 | Mean Train Rewards:   102.6 | Mean Test Rewards:   145.1 |
Reached reward test threshold in 2000 episodes
Reached reward test threshold in 2001 episodes
Reached reward test threshold in 2002 episodes
Reached reward train threshold in 2006 episodes
Reached reward test threshold in 2006 episodes
Reached reward train threshold in 2007 episodes
| Episode: 2010 | Mean Train Rewards:    98.6 | Mean Test Rewards:   143.7 |
Reached reward test threshold in 2010 episodes
Reached reward test threshold in 2013 episodes
Reached reward test threshold in 2015 episodes
Reached reward test threshold in 2016 episodes
Reached reward test threshold in 2017 episodes
| Episode: 2020 | Mean Train Rewards:    92.1 | Mean Test Rewards:   165.7 |
Reached reward test threshold in 2020 episodes
Reached reward test threshold in 2021 episodes
Reached reward train threshold in 2022 episodes
Reached reward test threshold in 2022 episodes
Reached reward train threshold in 2024 episodes
Reached reward test threshold in 2024 episodes
```

```
Reached reward train threshold in 2025 episodes
Reached reward test threshold in 2025 episodes
| Episode: 2030 | Mean Train Rewards:   111.8 | Mean Test Rewards:   171.1 |
Reached reward test threshold in 2030 episodes
Reached reward test threshold in 2032 episodes
Reached reward train threshold in 2034 episodes
Reached reward test threshold in 2035 episodes
Reached reward test threshold in 2037 episodes
Reached reward test threshold in 2038 episodes
| Episode: 2040 | Mean Train Rewards:    91.6 | Mean Test Rewards:   143.1 |
Reached reward train threshold in 2042 episodes
Reached reward test threshold in 2043 episodes
Reached reward train threshold in 2045 episodes
Reached reward test threshold in 2045 episodes
Reached reward test threshold in 2046 episodes
Reached reward test threshold in 2048 episodes
Reached reward train threshold in 2049 episodes
| Episode: 2050 | Mean Train Rewards:    89.5 | Mean Test Rewards:   109.2 |
Reached reward train threshold in 2051 episodes
Reached reward test threshold in 2051 episodes
Reached reward test threshold in 2054 episodes
Reached reward train threshold in 2058 episodes
Reached reward test threshold in 2059 episodes
| Episode: 2060 | Mean Train Rewards:   130.3 | Mean Test Rewards:   132.4 |
Reached reward test threshold in 2060 episodes
Reached reward test threshold in 2064 episodes
Reached reward test threshold in 2067 episodes
Reached reward test threshold in 2069 episodes
| Episode: 2070 | Mean Train Rewards:   102.9 | Mean Test Rewards:   149.4 |
Reached reward test threshold in 2070 episodes
Reached reward train threshold in 2071 episodes
Reached reward test threshold in 2074 episodes
Reached reward test threshold in 2078 episodes
| Episode: 2080 | Mean Train Rewards:    86.9 | Mean Test Rewards:   148.1 |
Reached reward train threshold in 2082 episodes
Reached reward train threshold in 2083 episodes
Reached reward test threshold in 2084 episodes
Reached reward test threshold in 2085 episodes
Reached reward test threshold in 2086 episodes
Reached reward train threshold in 2087 episodes
| Episode: 2090 | Mean Train Rewards:   104.0 | Mean Test Rewards:   139.5 |
Reached reward test threshold in 2091 episodes
Reached reward test threshold in 2093 episodes
Reached reward test threshold in 2094 episodes
Reached reward train threshold in 2096 episodes
Reached reward test threshold in 2096 episodes
Reached reward test threshold in 2097 episodes
| Episode: 2100 | Mean Train Rewards:   127.3 | Mean Test Rewards:   134.7 |
Reached reward train threshold in 2102 episodes
Reached reward train threshold in 2105 episodes
Reached reward test threshold in 2107 episodes
Reached reward train threshold in 2109 episodes
Reached reward test threshold in 2109 episodes
| Episode: 2110 | Mean Train Rewards:   123.9 | Mean Test Rewards:   128.3 |
Reached reward train threshold in 2110 episodes
Reached reward test threshold in 2110 episodes
Reached reward test threshold in 2111 episodes
Reached reward train threshold in 2112 episodes
Reached reward test threshold in 2112 episodes
Reached reward train threshold in 2116 episodes
Reached reward test threshold in 2118 episodes
Reached reward test threshold in 2119 episodes
| Episode: 2120 | Mean Train Rewards:   124.3 | Mean Test Rewards:   115.1 |
Reached reward train threshold in 2129 episodes
| Episode: 2130 | Mean Train Rewards:    99.9 | Mean Test Rewards:   101.2 |
Reached reward test threshold in 2130 episodes
Reached reward train threshold in 2131 episodes
Reached reward test threshold in 2138 episodes
| Episode: 2140 | Mean Train Rewards:   106.8 | Mean Test Rewards:    83.8 |
```

```
Reached reward test threshold in 2141 episodes
Reached reward train threshold in 2144 episodes
Reached reward train threshold in 2146 episodes
Reached reward test threshold in 2147 episodes
Reached reward test threshold in 2148 episodes
| Episode: 2150 | Mean Train Rewards:   137.2 | Mean Test Rewards:    99.2 |
Reached reward train threshold in 2151 episodes
Reached reward test threshold in 2154 episodes
Reached reward train threshold in 2157 episodes
Reached reward train threshold in 2158 episodes
Reached reward train threshold in 2159 episodes
| Episode: 2160 | Mean Train Rewards:   137.4 | Mean Test Rewards:   109.8 |
Reached reward test threshold in 2160 episodes
Reached reward train threshold in 2161 episodes
Reached reward test threshold in 2161 episodes
Reached reward train threshold in 2162 episodes
Reached reward test threshold in 2163 episodes
Reached reward train threshold in 2167 episodes
Reached reward test threshold in 2167 episodes
Reached reward train threshold in 2168 episodes
Reached reward test threshold in 2168 episodes
Reached reward test threshold in 2169 episodes
| Episode: 2170 | Mean Train Rewards:   158.8 | Mean Test Rewards:   133.0 |
Reached reward test threshold in 2171 episodes
Reached reward test threshold in 2174 episodes
Reached reward test threshold in 2175 episodes
Reached reward test threshold in 2176 episodes
Reached reward train threshold in 2177 episodes
Reached reward train threshold in 2179 episodes
Reached reward test threshold in 2179 episodes
| Episode: 2180 | Mean Train Rewards:   155.2 | Mean Test Rewards:   131.6 |
Reached reward train threshold in 2182 episodes
Reached reward test threshold in 2184 episodes
Reached reward train threshold in 2185 episodes
Reached reward test threshold in 2186 episodes
Reached reward train threshold in 2187 episodes
Reached reward test threshold in 2187 episodes
Reached reward train threshold in 2188 episodes
| Episode: 2190 | Mean Train Rewards:   145.7 | Mean Test Rewards:   157.3 |
Reached reward test threshold in 2193 episodes
Reached reward train threshold in 2194 episodes
Reached reward train threshold in 2195 episodes
Reached reward test threshold in 2195 episodes
Reached reward test threshold in 2196 episodes
Reached reward train threshold in 2199 episodes
Reached reward test threshold in 2199 episodes
| Episode: 2200 | Mean Train Rewards:   143.2 | Mean Test Rewards:   135.0 |
Reached reward train threshold in 2200 episodes
Reached reward test threshold in 2201 episodes
Reached reward train threshold in 2202 episodes
Reached reward train threshold in 2203 episodes
Reached reward test threshold in 2203 episodes
Reached reward train threshold in 2206 episodes
Reached reward train threshold in 2207 episodes
Reached reward test threshold in 2208 episodes
| Episode: 2210 | Mean Train Rewards:   167.2 | Mean Test Rewards:   149.3 |
Reached reward train threshold in 2210 episodes
Reached reward test threshold in 2210 episodes
Reached reward train threshold in 2211 episodes
Reached reward train threshold in 2212 episodes
Reached reward test threshold in 2212 episodes
Reached reward train threshold in 2216 episodes
Reached reward test threshold in 2216 episodes
Reached reward test threshold in 2218 episodes
| Episode: 2220 | Mean Train Rewards:   161.3 | Mean Test Rewards:   147.3 |
Reached reward train threshold in 2220 episodes
Reached reward test threshold in 2221 episodes
Reached reward train threshold in 2222 episodes
Reached reward train threshold in 2223 episodes
```

```
Reached reward train threshold in 2224 episodes
Reached reward test threshold in 2225 episodes
Reached reward test threshold in 2227 episodes
| Episode: 2230 | Mean Train Rewards:   158.9 | Mean Test Rewards:   157.6 |
Reached reward train threshold in 2230 episodes
Reached reward train threshold in 2231 episodes
Reached reward train threshold in 2232 episodes
Reached reward train threshold in 2233 episodes
Reached reward test threshold in 2233 episodes
Reached reward train threshold in 2234 episodes
Reached reward test threshold in 2234 episodes
Reached reward train threshold in 2235 episodes
Reached reward train threshold in 2236 episodes
Reached reward test threshold in 2236 episodes
Reached reward train threshold in 2238 episodes
Reached reward test threshold in 2239 episodes
| Episode: 2240 | Mean Train Rewards:   181.1 | Mean Test Rewards:   152.1 |
Reached reward train threshold in 2240 episodes
Reached reward test threshold in 2243 episodes
Reached reward test threshold in 2244 episodes
Reached reward train threshold in 2245 episodes
Reached reward train threshold in 2247 episodes
Reached reward test threshold in 2249 episodes
```



```
| Episode: 2250 | Mean Train Rewards:   171.7 | Mean Test Rewards:   148.5 |
Reached reward train threshold in 2250 episodes
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

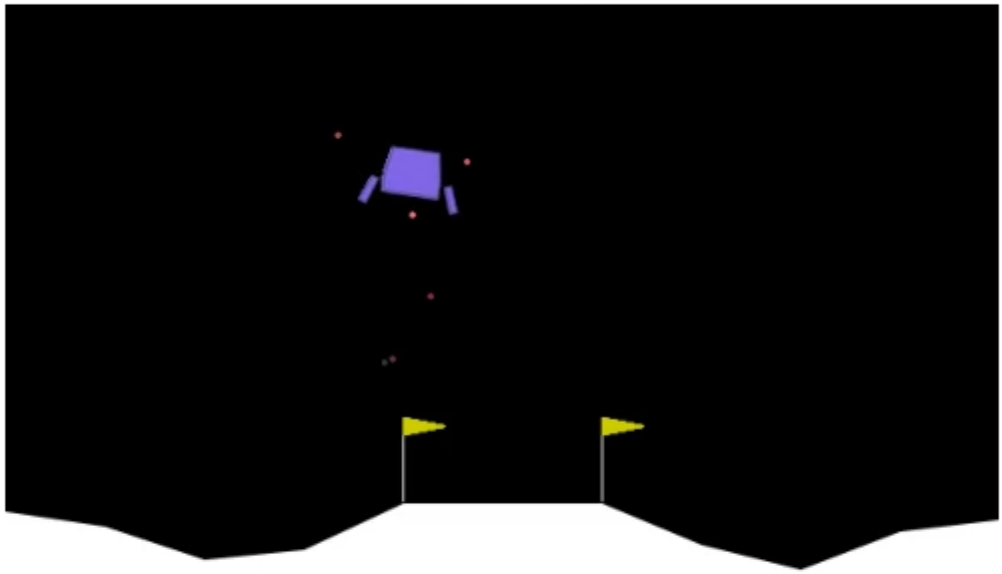
```
Reached reward test threshold in 2252 episodes
Reached reward test threshold in 2253 episodes
Reached reward test threshold in 2255 episodes
Reached reward train threshold in 2257 episodes
Reached reward train threshold in 2258 episodes
Reached reward train threshold in 2259 episodes
| Episode: 2260 | Mean Train Rewards:   154.5 | Mean Test Rewards:   154.6 |
Reached reward train threshold in 2261 episodes
Reached reward test threshold in 2261 episodes
Reached reward test threshold in 2262 episodes
Reached reward test threshold in 2264 episodes
Reached reward train threshold in 2265 episodes
Reached reward test threshold in 2265 episodes
Reached reward train threshold in 2266 episodes
Reached reward test threshold in 2268 episodes
Reached reward train threshold in 2269 episodes
| Episode: 2270 | Mean Train Rewards:   161.0 | Mean Test Rewards:   159.2 |
Reached reward train threshold in 2270 episodes
Reached reward train threshold in 2272 episodes
Reached reward test threshold in 2272 episodes
Reached reward test threshold in 2273 episodes
Reached reward train threshold in 2275 episodes
Reached reward test threshold in 2275 episodes
Reached reward test threshold in 2277 episodes
Reached reward test threshold in 2279 episodes
| Episode: 2280 | Mean Train Rewards:   165.3 | Mean Test Rewards:   171.7 |
Reached reward test threshold in 2280 episodes
Reached reward test threshold in 2282 episodes
Reached reward train threshold in 2283 episodes
Reached reward test threshold in 2284 episodes
Reached reward test threshold in 2289 episodes
| Episode: 2290 | Mean Train Rewards:   145.9 | Mean Test Rewards:   175.2 |
Reached reward test threshold in 2290 episodes
Reached reward test threshold in 2292 episodes
Reached reward test threshold in 2293 episodes
Reached reward test threshold in 2294 episodes
Reached reward test threshold in 2296 episodes
Reached reward train threshold in 2297 episodes
Reached reward train threshold in 2298 episodes
| Episode: 2300 | Mean Train Rewards:   134.6 | Mean Test Rewards:   167.3 |
Reached reward train threshold in 2304 episodes
| Episode: 2310 | Mean Train Rewards:   131.0 | Mean Test Rewards:   115.5 |
Reached reward train threshold in 2313 episodes
Reached reward train threshold in 2314 episodes
| Episode: 2320 | Mean Train Rewards:   135.1 | Mean Test Rewards:    53.4 |
| Episode: 2330 | Mean Train Rewards:   137.1 | Mean Test Rewards:    20.7 |
| Episode: 2340 | Mean Train Rewards:   130.7 | Mean Test Rewards:     6.0 |
Reached reward train threshold in 2342 episodes
| Episode: 2350 | Mean Train Rewards:   134.9 | Mean Test Rewards:     1.1 |
| Episode: 2360 | Mean Train Rewards:   133.2 | Mean Test Rewards:    13.8 |
Reached reward train threshold in 2367 episodes
| Episode: 2370 | Mean Train Rewards:   137.1 | Mean Test Rewards:    61.4 |
| Episode: 2380 | Mean Train Rewards:   136.5 | Mean Test Rewards:   114.7 |
Reached reward train threshold in 2381 episodes
Reached reward test threshold in 2388 episodes
| Episode: 2390 | Mean Train Rewards:   137.3 | Mean Test Rewards:   135.7 |
Reached reward test threshold in 2390 episodes
Reached reward test threshold in 2391 episodes
Reached reward test threshold in 2392 episodes
Reached reward test threshold in 2393 episodes
Reached reward test threshold in 2394 episodes
Reached reward test threshold in 2395 episodes
Reached reward test threshold in 2396 episodes
Reached reward test threshold in 2397 episodes
Reached reward test threshold in 2399 episodes
| Episode: 2400 | Mean Train Rewards:   135.7 | Mean Test Rewards:   173.8 |
Reached reward test threshold in 2400 episodes
Reached reward test threshold in 2401 episodes
Reached reward test threshold in 2402 episodes
```

```
Reached reward test threshold in 2403 episodes
Reached reward test threshold in 2404 episodes
Reached reward train threshold in 2405 episodes
Reached reward test threshold in 2405 episodes
Reached reward test threshold in 2406 episodes
Reached reward test threshold in 2409 episodes
| Episode: 2410 | Mean Train Rewards:   141.5 | Mean Test Rewards:   211.7 |
Reached reward train threshold in 2410 episodes
Reached reward test threshold in 2410 episodes
| Episode: 2420 | Mean Train Rewards:   144.8 | Mean Test Rewards:   173.7 |
Reached reward train threshold in 2423 episodes
Reached reward train threshold in 2426 episodes
Reached reward test threshold in 2427 episodes
| Episode: 2430 | Mean Train Rewards:   142.2 | Mean Test Rewards:   132.3 |
Reached reward train threshold in 2431 episodes
Reached reward test threshold in 2438 episodes
Reached reward test threshold in 2439 episodes
| Episode: 2440 | Mean Train Rewards:   144.5 | Mean Test Rewards:   130.6 |
Reached reward test threshold in 2442 episodes
Reached reward train threshold in 2443 episodes
Reached reward test threshold in 2443 episodes
Reached reward train threshold in 2446 episodes
Reached reward train threshold in 2447 episodes
Reached reward test threshold in 2448 episodes
Reached reward test threshold in 2449 episodes
| Episode: 2450 | Mean Train Rewards:   151.3 | Mean Test Rewards:   154.2 |
Reached reward test threshold in 2452 episodes
Reached reward test threshold in 2453 episodes
Reached reward test threshold in 2455 episodes
Reached reward test threshold in 2456 episodes
Reached reward test threshold in 2457 episodes
Reached reward test threshold in 2459 episodes
| Episode: 2460 | Mean Train Rewards:   133.3 | Mean Test Rewards:   177.1 |
Reached reward test threshold in 2462 episodes
Reached reward test threshold in 2464 episodes
Reached reward test threshold in 2466 episodes
Reached reward test threshold in 2467 episodes
Reached reward test threshold in 2468 episodes
| Episode: 2470 | Mean Train Rewards:   114.1 | Mean Test Rewards:   190.3 |
Reached reward test threshold in 2470 episodes
Reached reward test threshold in 2474 episodes
Reached reward test threshold in 2476 episodes
Reached reward test threshold in 2477 episodes
Reached reward test threshold in 2478 episodes
Reached reward test threshold in 2479 episodes
| Episode: 2480 | Mean Train Rewards:    91.4 | Mean Test Rewards:   182.0 |
Reached reward test threshold in 2481 episodes
Reached reward test threshold in 2485 episodes
Reached reward test threshold in 2486 episodes
Reached reward test threshold in 2487 episodes
Reached reward test threshold in 2489 episodes
| Episode: 2490 | Mean Train Rewards:    89.0 | Mean Test Rewards:   185.9 |
Reached reward test threshold in 2491 episodes
Reached reward test threshold in 2493 episodes
Reached reward test threshold in 2494 episodes
Reached reward train threshold in 2496 episodes
Reached reward test threshold in 2496 episodes
Reached reward test threshold in 2498 episodes
| Episode: 2500 | Mean Train Rewards:   110.6 | Mean Test Rewards:   185.2 |
Reached reward train threshold in 2501 episodes
Reached reward test threshold in 2502 episodes
Reached reward train threshold in 2508 episodes
| Episode: 2510 | Mean Train Rewards:   135.9 | Mean Test Rewards:   163.0 |
Reached reward test threshold in 2517 episodes
| Episode: 2520 | Mean Train Rewards:   136.3 | Mean Test Rewards:   142.2 |
Reached reward train threshold in 2521 episodes
Reached reward test threshold in 2524 episodes
Reached reward test threshold in 2528 episodes
Reached reward test threshold in 2529 episodes
```

```
| Episode: 2530 | Mean Train Rewards:   128.6 | Mean Test Rewards:   142.3 |
Reached reward test threshold in 2531 episodes
Reached reward test threshold in 2533 episodes
Reached reward test threshold in 2538 episodes
| Episode: 2540 | Mean Train Rewards:   126.8 | Mean Test Rewards:   163.8 |
Reached reward test threshold in 2540 episodes
Reached reward test threshold in 2541 episodes
| Episode: 2550 | Mean Train Rewards:   131.9 | Mean Test Rewards:   148.6 |
Reached reward test threshold in 2554 episodes
| Episode: 2560 | Mean Train Rewards:   141.4 | Mean Test Rewards:   133.6 |
| Episode: 2570 | Mean Train Rewards:   141.5 | Mean Test Rewards:   130.1 |
Reached reward test threshold in 2570 episodes
Reached reward train threshold in 2572 episodes
| Episode: 2580 | Mean Train Rewards:   145.3 | Mean Test Rewards:   129.4 |
Reached reward test threshold in 2587 episodes
| Episode: 2590 | Mean Train Rewards:   141.3 | Mean Test Rewards:   137.4 |
Reached reward train threshold in 2591 episodes
Reached reward test threshold in 2593 episodes
Reached reward test threshold in 2594 episodes
Reached reward test threshold in 2595 episodes
Reached reward test threshold in 2596 episodes
Reached reward test threshold in 2597 episodes
Reached reward test threshold in 2598 episodes
Reached reward test threshold in 2599 episodes
| Episode: 2600 | Mean Train Rewards:   130.1 | Mean Test Rewards:   171.2 |
Reached reward test threshold in 2600 episodes
Reached reward test threshold in 2601 episodes
Reached reward test threshold in 2602 episodes
Reached reward test threshold in 2603 episodes
Reached reward test threshold in 2604 episodes
Reached reward test threshold in 2605 episodes
Reached reward test threshold in 2607 episodes
Reached reward train threshold in 2609 episodes
Reached reward test threshold in 2609 episodes
| Episode: 2610 | Mean Train Rewards:   110.9 | Mean Test Rewards:   202.4 |
Reached reward test threshold in 2610 episodes
Reached reward test threshold in 2611 episodes
Reached reward train threshold in 2612 episodes
Reached reward test threshold in 2612 episodes
Reached reward test threshold in 2613 episodes
Reached reward train threshold in 2615 episodes
Reached reward test threshold in 2615 episodes
Reached reward train threshold in 2616 episodes
Reached reward test threshold in 2616 episodes
Reached reward test threshold in 2618 episodes
| Episode: 2620 | Mean Train Rewards:   124.2 | Mean Test Rewards:   216.8 |
Reached reward train threshold in 2620 episodes
Reached reward test threshold in 2620 episodes
Reached reward test threshold in 2621 episodes
Reached reward test threshold in 2622 episodes
Reached reward test threshold in 2623 episodes
Reached reward train threshold in 2625 episodes
Reached reward test threshold in 2627 episodes
Reached reward test threshold in 2628 episodes
Reached reward test threshold in 2629 episodes
| Episode: 2630 | Mean Train Rewards:   120.1 | Mean Test Rewards:   192.5 |
Reached reward test threshold in 2631 episodes
Reached reward test threshold in 2632 episodes
Reached reward train threshold in 2634 episodes
| Episode: 2640 | Mean Train Rewards:   126.7 | Mean Test Rewards:   182.5 |
Reached reward train threshold in 2640 episodes
Reached reward test threshold in 2640 episodes
Reached reward test threshold in 2646 episodes
Reached reward test threshold in 2648 episodes
Reached reward test threshold in 2649 episodes
| Episode: 2650 | Mean Train Rewards:   129.9 | Mean Test Rewards:   173.9 |
Reached reward train threshold in 2650 episodes
Reached reward train threshold in 2659 episodes
| Episode: 2660 | Mean Train Rewards:   128.2 | Mean Test Rewards:   157.4 |
```

```
Reached reward train threshold in 2667 episodes
| Episode: 2670 | Mean Train Rewards:   130.1 | Mean Test Rewards:   145.5 |
| Episode: 2680 | Mean Train Rewards:   136.6 | Mean Test Rewards:   128.6 |
Reached reward test threshold in 2683 episodes
| Episode: 2690 | Mean Train Rewards:   131.0 | Mean Test Rewards:   131.7 |
Reached reward train threshold in 2692 episodes
| Episode: 2700 | Mean Train Rewards:   127.4 | Mean Test Rewards:   136.9 |
| Episode: 2710 | Mean Train Rewards:   126.9 | Mean Test Rewards:   131.4 |
| Episode: 2720 | Mean Train Rewards:   130.4 | Mean Test Rewards:   128.7 |
| Episode: 2730 | Mean Train Rewards:   128.8 | Mean Test Rewards:   130.3 |
Reached reward test threshold in 2738 episodes
| Episode: 2740 | Mean Train Rewards:   123.7 | Mean Test Rewards:   132.4 |
Reached reward train threshold in 2748 episodes
Reached reward test threshold in 2749 episodes
| Episode: 2750 | Mean Train Rewards:   114.8 | Mean Test Rewards:   144.2 |
Reached reward test threshold in 2753 episodes
Reached reward train threshold in 2755 episodes
Reached reward test threshold in 2755 episodes
| Episode: 2760 | Mean Train Rewards:   132.1 | Mean Test Rewards:   157.2 |
| Episode: 2770 | Mean Train Rewards:   136.9 | Mean Test Rewards:   148.2 |
Reached reward test threshold in 2771 episodes
Reached reward test threshold in 2773 episodes
Reached reward train threshold in 2775 episodes
Reached reward test threshold in 2778 episodes
Reached reward test threshold in 2779 episodes
| Episode: 2780 | Mean Train Rewards:   137.7 | Mean Test Rewards:   138.9 |
Reached reward test threshold in 2784 episodes
Reached reward test threshold in 2787 episodes
| Episode: 2790 | Mean Train Rewards:   132.8 | Mean Test Rewards:   149.0 |
Reached reward test threshold in 2791 episodes
Reached reward train threshold in 2793 episodes
Reached reward test threshold in 2793 episodes
Reached reward train threshold in 2794 episodes
Reached reward test threshold in 2795 episodes
Reached reward test threshold in 2797 episodes
Reached reward test threshold in 2798 episodes
Reached reward test threshold in 2799 episodes
| Episode: 2800 | Mean Train Rewards:   133.6 | Mean Test Rewards:   160.5 |
Reached reward train threshold in 2801 episodes
Reached reward train threshold in 2802 episodes
Reached reward test threshold in 2804 episodes
Reached reward test threshold in 2806 episodes
Reached reward test threshold in 2808 episodes
| Episode: 2810 | Mean Train Rewards:   128.7 | Mean Test Rewards:   143.9 |
Reached reward test threshold in 2817 episodes
Reached reward test threshold in 2819 episodes
| Episode: 2820 | Mean Train Rewards:    88.7 | Mean Test Rewards:   126.3 |
Reached reward test threshold in 2820 episodes
Reached reward test threshold in 2821 episodes
Reached reward test threshold in 2824 episodes
Reached reward test threshold in 2825 episodes
Reached reward test threshold in 2826 episodes
Reached reward test threshold in 2827 episodes
Reached reward test threshold in 2828 episodes
Reached reward test threshold in 2829 episodes
| Episode: 2830 | Mean Train Rewards:    89.6 | Mean Test Rewards:   151.5 |
Reached reward test threshold in 2830 episodes
Reached reward test threshold in 2831 episodes
Reached reward train threshold in 2833 episodes
Reached reward test threshold in 2833 episodes
Reached reward train threshold in 2835 episodes
Reached reward test threshold in 2837 episodes
Reached reward test threshold in 2839 episodes
| Episode: 2840 | Mean Train Rewards:   113.8 | Mean Test Rewards:   187.3 |
Reached reward test threshold in 2840 episodes
Reached reward test threshold in 2848 episodes
| Episode: 2850 | Mean Train Rewards:   133.5 | Mean Test Rewards:   173.4 |
Reached reward train threshold in 2855 episodes
Reached reward test threshold in 2859 episodes
```

```
| Episode: 2860 | Mean Train Rewards:   121.1 | Mean Test Rewards:   145.7 |
Reached reward train threshold in 2862 episodes
Reached reward test threshold in 2865 episodes
Reached reward train threshold in 2866 episodes
| Episode: 2870 | Mean Train Rewards:   136.7 | Mean Test Rewards:   148.4 |
| Episode: 2880 | Mean Train Rewards:   140.8 | Mean Test Rewards:   141.9 |
| Episode: 2890 | Mean Train Rewards:   140.2 | Mean Test Rewards:   129.1 |
Reached reward train threshold in 2894 episodes
| Episode: 2900 | Mean Train Rewards:   136.4 | Mean Test Rewards:   126.4 |
| Episode: 2910 | Mean Train Rewards:   139.2 | Mean Test Rewards:   123.7 |
| Episode: 2920 | Mean Train Rewards:   133.7 | Mean Test Rewards:   130.3 |
| Episode: 2930 | Mean Train Rewards:   136.4 | Mean Test Rewards:   123.4 |
Reached reward test threshold in 2938 episodes
| Episode: 2940 | Mean Train Rewards:   137.1 | Mean Test Rewards:   128.3 |
| Episode: 2950 | Mean Train Rewards:   150.5 | Mean Test Rewards:   128.3 |
| Episode: 2960 | Mean Train Rewards:   143.5 | Mean Test Rewards:   126.5 |
Reached reward train threshold in 2960 episodes
Reached reward train threshold in 2961 episodes
Reached reward train threshold in 2969 episodes
| Episode: 2970 | Mean Train Rewards:   141.9 | Mean Test Rewards:   117.5 |
Reached reward train threshold in 2975 episodes
| Episode: 2980 | Mean Train Rewards:   137.0 | Mean Test Rewards:   113.0 |
| Episode: 2990 | Mean Train Rewards:   128.7 | Mean Test Rewards:   108.3 |
```



```
| Episode: 3000 | Mean Train Rewards:   126.0 | Mean Test Rewards:   113.6 |
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

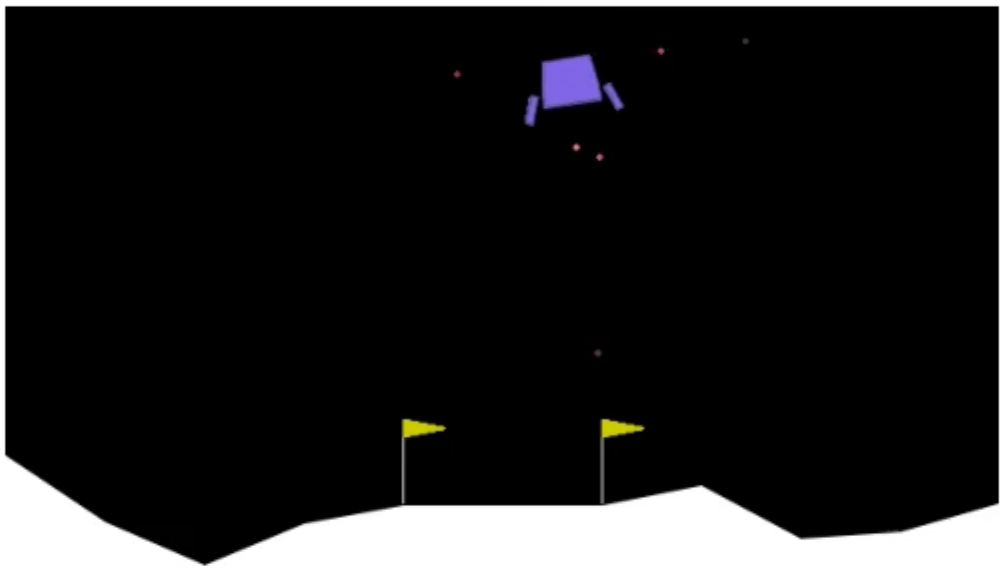
```
| Episode: 3010 | Mean Train Rewards:   138.8 | Mean Test Rewards:   105.3 |
Reached reward train threshold in 3011 episodes
Reached reward test threshold in 3011 episodes
Reached reward train threshold in 3015 episodes
Reached reward test threshold in 3017 episodes
Reached reward train threshold in 3019 episodes
| Episode: 3020 | Mean Train Rewards:   152.4 | Mean Test Rewards:   111.1 |
Reached reward train threshold in 3028 episodes
Reached reward train threshold in 3029 episodes
Reached reward test threshold in 3029 episodes
| Episode: 3030 | Mean Train Rewards:   137.8 | Mean Test Rewards:   114.5 |
Reached reward train threshold in 3037 episodes
| Episode: 3040 | Mean Train Rewards:   102.2 | Mean Test Rewards:   121.3 |
| Episode: 3050 | Mean Train Rewards:    88.4 | Mean Test Rewards:   118.7 |
Reached reward train threshold in 3053 episodes
Reached reward test threshold in 3054 episodes
| Episode: 3060 | Mean Train Rewards:    91.4 | Mean Test Rewards:   120.4 |
Reached reward train threshold in 3064 episodes
Reached reward train threshold in 3069 episodes
| Episode: 3070 | Mean Train Rewards:   121.4 | Mean Test Rewards:   116.8 |
Reached reward train threshold in 3071 episodes
Reached reward test threshold in 3071 episodes
Reached reward train threshold in 3074 episodes
Reached reward train threshold in 3077 episodes
Reached reward test threshold in 3078 episodes
Reached reward test threshold in 3079 episodes
| Episode: 3080 | Mean Train Rewards:   124.4 | Mean Test Rewards:   118.4 |
Reached reward test threshold in 3082 episodes
Reached reward test threshold in 3084 episodes
Reached reward test threshold in 3086 episodes
Reached reward train threshold in 3087 episodes
Reached reward test threshold in 3087 episodes
Reached reward test threshold in 3089 episodes
| Episode: 3090 | Mean Train Rewards:   134.8 | Mean Test Rewards:   149.4 |
Reached reward test threshold in 3090 episodes
Reached reward train threshold in 3091 episodes
Reached reward test threshold in 3092 episodes
Reached reward train threshold in 3093 episodes
Reached reward test threshold in 3093 episodes
Reached reward test threshold in 3094 episodes
Reached reward test threshold in 3095 episodes
Reached reward test threshold in 3096 episodes
| Episode: 3100 | Mean Train Rewards:   147.5 | Mean Test Rewards:   173.0 |
Reached reward test threshold in 3105 episodes
Reached reward test threshold in 3109 episodes
| Episode: 3110 | Mean Train Rewards:   148.2 | Mean Test Rewards:   162.2 |
Reached reward test threshold in 3113 episodes
| Episode: 3120 | Mean Train Rewards:   137.4 | Mean Test Rewards:   145.8 |
| Episode: 3130 | Mean Train Rewards:   148.3 | Mean Test Rewards:   135.1 |
| Episode: 3140 | Mean Train Rewards:   144.1 | Mean Test Rewards:   118.1 |
Reached reward test threshold in 3144 episodes
Reached reward test threshold in 3148 episodes
| Episode: 3150 | Mean Train Rewards:   137.4 | Mean Test Rewards:   130.4 |
Reached reward train threshold in 3150 episodes
Reached reward test threshold in 3154 episodes
Reached reward train threshold in 3157 episodes
| Episode: 3160 | Mean Train Rewards:   136.1 | Mean Test Rewards:   149.5 |
Reached reward test threshold in 3164 episodes
Reached reward test threshold in 3167 episodes
| Episode: 3170 | Mean Train Rewards:   128.7 | Mean Test Rewards:   151.6 |
Reached reward train threshold in 3173 episodes
Reached reward test threshold in 3176 episodes
Reached reward test threshold in 3177 episodes
Reached reward test threshold in 3178 episodes
Reached reward test threshold in 3179 episodes
| Episode: 3180 | Mean Train Rewards:   125.0 | Mean Test Rewards:   150.8 |
Reached reward test threshold in 3181 episodes
Reached reward test threshold in 3184 episodes
Reached reward train threshold in 3185 episodes
```

```
Reached reward test threshold in 3185 episodes
| Episode: 3190 | Mean Train Rewards:   124.6 | Mean Test Rewards:   154.4 |
Reached reward test threshold in 3191 episodes
Reached reward train threshold in 3196 episodes
Reached reward test threshold in 3196 episodes
Reached reward train threshold in 3198 episodes
Reached reward test threshold in 3198 episodes
Reached reward test threshold in 3199 episodes
| Episode: 3200 | Mean Train Rewards:   122.0 | Mean Test Rewards:   179.0 |
Reached reward test threshold in 3200 episodes
Reached reward train threshold in 3205 episodes
Reached reward test threshold in 3209 episodes
| Episode: 3210 | Mean Train Rewards:   110.3 | Mean Test Rewards:   154.4 |
Reached reward test threshold in 3212 episodes
| Episode: 3220 | Mean Train Rewards:   118.9 | Mean Test Rewards:   156.6 |
Reached reward test threshold in 3221 episodes
Reached reward test threshold in 3222 episodes
Reached reward train threshold in 3223 episodes
Reached reward test threshold in 3223 episodes
Reached reward train threshold in 3226 episodes
Reached reward test threshold in 3227 episodes
| Episode: 3230 | Mean Train Rewards:   123.1 | Mean Test Rewards:   155.8 |
Reached reward test threshold in 3233 episodes
| Episode: 3240 | Mean Train Rewards:   130.6 | Mean Test Rewards:   149.9 |
| Episode: 3250 | Mean Train Rewards:   126.0 | Mean Test Rewards:   125.8 |
| Episode: 3260 | Mean Train Rewards:   136.2 | Mean Test Rewards:   113.2 |
| Episode: 3270 | Mean Train Rewards:   134.1 | Mean Test Rewards:   108.5 |
Reached reward train threshold in 3271 episodes
Reached reward test threshold in 3273 episodes
| Episode: 3280 | Mean Train Rewards:   133.8 | Mean Test Rewards:   116.4 |
Reached reward train threshold in 3283 episodes
| Episode: 3290 | Mean Train Rewards:   131.9 | Mean Test Rewards:   115.8 |
| Episode: 3300 | Mean Train Rewards:   134.6 | Mean Test Rewards:   103.0 |
Reached reward test threshold in 3302 episodes
| Episode: 3310 | Mean Train Rewards:   131.9 | Mean Test Rewards:   114.1 |
| Episode: 3320 | Mean Train Rewards:   138.0 | Mean Test Rewards:   114.3 |
| Episode: 3330 | Mean Train Rewards:   132.0 | Mean Test Rewards:   103.8 |
| Episode: 3340 | Mean Train Rewards:   133.9 | Mean Test Rewards:   109.7 |
| Episode: 3350 | Mean Train Rewards:   139.6 | Mean Test Rewards:   114.6 |
Reached reward test threshold in 3352 episodes
Reached reward test threshold in 3359 episodes
| Episode: 3360 | Mean Train Rewards:   130.0 | Mean Test Rewards:   133.4 |
Reached reward train threshold in 3360 episodes
Reached reward train threshold in 3366 episodes
Reached reward test threshold in 3366 episodes
| Episode: 3370 | Mean Train Rewards:   114.7 | Mean Test Rewards:   134.9 |
Reached reward train threshold in 3371 episodes
| Episode: 3380 | Mean Train Rewards:   125.0 | Mean Test Rewards:   131.7 |
Reached reward test threshold in 3389 episodes
| Episode: 3390 | Mean Train Rewards:   136.0 | Mean Test Rewards:   131.0 |
Reached reward train threshold in 3393 episodes
| Episode: 3400 | Mean Train Rewards:   127.6 | Mean Test Rewards:   128.4 |
Reached reward train threshold in 3404 episodes
Reached reward train threshold in 3409 episodes
| Episode: 3410 | Mean Train Rewards:   133.6 | Mean Test Rewards:   126.2 |
Reached reward test threshold in 3413 episodes
Reached reward train threshold in 3418 episodes
Reached reward test threshold in 3419 episodes
| Episode: 3420 | Mean Train Rewards:   129.6 | Mean Test Rewards:   134.9 |
Reached reward train threshold in 3422 episodes
| Episode: 3430 | Mean Train Rewards:   137.1 | Mean Test Rewards:   135.1 |
Reached reward train threshold in 3435 episodes
Reached reward train threshold in 3436 episodes
| Episode: 3440 | Mean Train Rewards:   123.4 | Mean Test Rewards:   126.0 |
| Episode: 3450 | Mean Train Rewards:   118.3 | Mean Test Rewards:   116.5 |
Reached reward train threshold in 3452 episodes
Reached reward test threshold in 3455 episodes
Reached reward test threshold in 3458 episodes
Reached reward test threshold in 3459 episodes
```

```
| Episode: 3460 | Mean Train Rewards:   120.9 | Mean Test Rewards:   130.1 |
Reached reward test threshold in 3461 episodes
Reached reward test threshold in 3462 episodes
Reached reward test threshold in 3464 episodes
Reached reward test threshold in 3465 episodes
Reached reward train threshold in 3466 episodes
Reached reward test threshold in 3466 episodes
| Episode: 3470 | Mean Train Rewards:   117.8 | Mean Test Rewards:   162.4 |
Reached reward test threshold in 3470 episodes
Reached reward train threshold in 3478 episodes
Reached reward test threshold in 3479 episodes
| Episode: 3480 | Mean Train Rewards:   115.6 | Mean Test Rewards:   166.3 |
Reached reward train threshold in 3484 episodes
Reached reward test threshold in 3488 episodes
Reached reward test threshold in 3489 episodes
| Episode: 3490 | Mean Train Rewards:   152.0 | Mean Test Rewards:   132.2 |
Reached reward train threshold in 3490 episodes
Reached reward train threshold in 3491 episodes
Reached reward train threshold in 3493 episodes
Reached reward train threshold in 3494 episodes
| Episode: 3500 | Mean Train Rewards:   168.5 | Mean Test Rewards:   121.3 |
Reached reward train threshold in 3500 episodes
Reached reward train threshold in 3501 episodes
| Episode: 3510 | Mean Train Rewards:   155.0 | Mean Test Rewards:   128.5 |
Reached reward test threshold in 3511 episodes
| Episode: 3520 | Mean Train Rewards:   124.1 | Mean Test Rewards:   127.4 |
Reached reward test threshold in 3524 episodes
Reached reward test threshold in 3526 episodes
| Episode: 3530 | Mean Train Rewards:   121.5 | Mean Test Rewards:   128.0 |
Reached reward train threshold in 3531 episodes
Reached reward test threshold in 3538 episodes
| Episode: 3540 | Mean Train Rewards:   130.2 | Mean Test Rewards:   129.5 |
Reached reward train threshold in 3540 episodes
Reached reward test threshold in 3542 episodes
Reached reward train threshold in 3544 episodes
Reached reward test threshold in 3544 episodes
| Episode: 3550 | Mean Train Rewards:   121.0 | Mean Test Rewards:   128.2 |
Reached reward test threshold in 3552 episodes
Reached reward test threshold in 3554 episodes
| Episode: 3560 | Mean Train Rewards:   124.0 | Mean Test Rewards:   137.8 |
Reached reward train threshold in 3560 episodes
Reached reward test threshold in 3561 episodes
Reached reward test threshold in 3563 episodes
Reached reward test threshold in 3564 episodes
Reached reward test threshold in 3568 episodes
Reached reward test threshold in 3569 episodes
| Episode: 3570 | Mean Train Rewards:   103.5 | Mean Test Rewards:   149.9 |
Reached reward test threshold in 3570 episodes
Reached reward test threshold in 3574 episodes
Reached reward test threshold in 3575 episodes
| Episode: 3580 | Mean Train Rewards:    91.3 | Mean Test Rewards:   150.6 |
Reached reward test threshold in 3581 episodes
Reached reward test threshold in 3584 episodes
Reached reward test threshold in 3585 episodes
Reached reward train threshold in 3586 episodes
Reached reward test threshold in 3586 episodes
Reached reward test threshold in 3587 episodes
Reached reward test threshold in 3588 episodes
Reached reward train threshold in 3589 episodes
| Episode: 3590 | Mean Train Rewards:    95.8 | Mean Test Rewards:   165.9 |
Reached reward test threshold in 3590 episodes
Reached reward train threshold in 3591 episodes
Reached reward test threshold in 3591 episodes
Reached reward train threshold in 3593 episodes
Reached reward test threshold in 3593 episodes
Reached reward test threshold in 3594 episodes
Reached reward train threshold in 3595 episodes
Reached reward train threshold in 3596 episodes
Reached reward test threshold in 3596 episodes
```

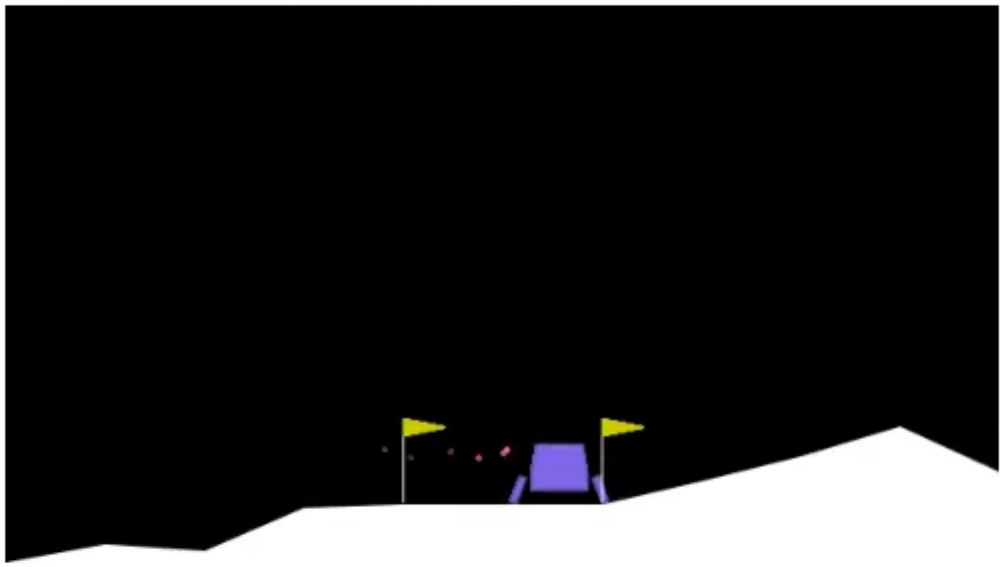
```
Reached reward train threshold in 3597 episodes
Reached reward test threshold in 3598 episodes
Reached reward train threshold in 3599 episodes
Reached reward test threshold in 3599 episodes
| Episode: 3600 | Mean Train Rewards:   131.9 | Mean Test Rewards:   169.4 |
Reached reward test threshold in 3601 episodes
Reached reward train threshold in 3602 episodes
Reached reward test threshold in 3604 episodes
Reached reward test threshold in 3605 episodes
Reached reward train threshold in 3608 episodes
Reached reward test threshold in 3608 episodes
Reached reward test threshold in 3609 episodes
| Episode: 3610 | Mean Train Rewards:   166.4 | Mean Test Rewards:   187.9 |
Reached reward test threshold in 3610 episodes
Reached reward train threshold in 3613 episodes
Reached reward test threshold in 3613 episodes
Reached reward train threshold in 3614 episodes
Reached reward test threshold in 3614 episodes
Reached reward train threshold in 3615 episodes
Reached reward train threshold in 3617 episodes
Reached reward test threshold in 3618 episodes
| Episode: 3620 | Mean Train Rewards:   159.5 | Mean Test Rewards:   175.2 |
Reached reward test threshold in 3620 episodes
Reached reward test threshold in 3621 episodes
Reached reward train threshold in 3622 episodes
Reached reward test threshold in 3622 episodes
Reached reward train threshold in 3623 episodes
Reached reward train threshold in 3625 episodes
Reached reward test threshold in 3627 episodes
Reached reward train threshold in 3628 episodes
Reached reward test threshold in 3628 episodes
| Episode: 3630 | Mean Train Rewards:   149.9 | Mean Test Rewards:   163.8 |
Reached reward test threshold in 3631 episodes
Reached reward train threshold in 3632 episodes
Reached reward train threshold in 3635 episodes
Reached reward train threshold in 3638 episodes
Reached reward test threshold in 3638 episodes
| Episode: 3640 | Mean Train Rewards:   141.8 | Mean Test Rewards:   156.9 |
Reached reward train threshold in 3640 episodes
Reached reward test threshold in 3640 episodes
Reached reward train threshold in 3643 episodes
Reached reward train threshold in 3644 episodes
Reached reward test threshold in 3644 episodes
Reached reward train threshold in 3648 episodes
Reached reward train threshold in 3649 episodes
| Episode: 3650 | Mean Train Rewards:   147.1 | Mean Test Rewards:   136.3 |
Reached reward train threshold in 3650 episodes
Reached reward test threshold in 3650 episodes
Reached reward train threshold in 3651 episodes
Reached reward test threshold in 3651 episodes
Reached reward train threshold in 3653 episodes
Reached reward test threshold in 3655 episodes
Reached reward test threshold in 3658 episodes
| Episode: 3660 | Mean Train Rewards:   161.9 | Mean Test Rewards:   137.7 |
Reached reward test threshold in 3660 episodes
Reached reward train threshold in 3667 episodes
| Episode: 3670 | Mean Train Rewards:   156.1 | Mean Test Rewards:   125.3 |
Reached reward train threshold in 3674 episodes
Reached reward train threshold in 3675 episodes
Reached reward train threshold in 3676 episodes
Reached reward train threshold in 3678 episodes
Reached reward test threshold in 3678 episodes
| Episode: 3680 | Mean Train Rewards:   148.7 | Mean Test Rewards:   117.8 |
Reached reward train threshold in 3681 episodes
Reached reward train threshold in 3684 episodes
Reached reward train threshold in 3685 episodes
Reached reward train threshold in 3687 episodes
Reached reward train threshold in 3688 episodes
Reached reward test threshold in 3689 episodes
```

```
| Episode: 3690 | Mean Train Rewards:   164.8 | Mean Test Rewards:   117.7 |
Reached reward train threshold in 3690 episodes
Reached reward train threshold in 3691 episodes
Reached reward test threshold in 3693 episodes
Reached reward train threshold in 3694 episodes
| Episode: 3700 | Mean Train Rewards:   164.6 | Mean Test Rewards:   118.1 |
Reached reward test threshold in 3705 episodes
Reached reward test threshold in 3706 episodes
Reached reward train threshold in 3709 episodes
| Episode: 3710 | Mean Train Rewards:   149.9 | Mean Test Rewards:   125.4 |
Reached reward train threshold in 3710 episodes
Reached reward train threshold in 3711 episodes
Reached reward train threshold in 3714 episodes
Reached reward train threshold in 3716 episodes
Reached reward test threshold in 3716 episodes
Reached reward train threshold in 3719 episodes
| Episode: 3720 | Mean Train Rewards:   145.7 | Mean Test Rewards:   137.0 |
Reached reward train threshold in 3720 episodes
Reached reward train threshold in 3724 episodes
Reached reward train threshold in 3728 episodes
| Episode: 3730 | Mean Train Rewards:   155.2 | Mean Test Rewards:   129.2 |
Reached reward train threshold in 3731 episodes
Reached reward train threshold in 3732 episodes
Reached reward train threshold in 3733 episodes
Reached reward train threshold in 3735 episodes
| Episode: 3740 | Mean Train Rewards:   154.2 | Mean Test Rewards:   100.8 |
```



```
| Episode: 3750 | Mean Train Rewards:   146.9 | Mean Test Rewards:    72.9 |
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

```
Reached reward train threshold in 3758 episodes
Reached reward train threshold in 3759 episodes
| Episode: 3760 | Mean Train Rewards:   134.7 | Mean Test Rewards:    79.3 |
Reached reward train threshold in 3768 episodes
| Episode: 3770 | Mean Train Rewards:   142.5 | Mean Test Rewards:    89.2 |
| Episode: 3780 | Mean Train Rewards:   140.8 | Mean Test Rewards:   105.2 |
| Episode: 3790 | Mean Train Rewards:   133.5 | Mean Test Rewards:   106.7 |
| Episode: 3800 | Mean Train Rewards:   137.3 | Mean Test Rewards:   109.4 |
Reached reward test threshold in 3803 episodes
| Episode: 3810 | Mean Train Rewards:   137.9 | Mean Test Rewards:   126.4 |
| Episode: 3820 | Mean Train Rewards:   139.0 | Mean Test Rewards:   129.6 |
Reached reward test threshold in 3827 episodes
| Episode: 3830 | Mean Train Rewards:   139.1 | Mean Test Rewards:   133.0 |
Reached reward test threshold in 3830 episodes
| Episode: 3840 | Mean Train Rewards:   137.2 | Mean Test Rewards:   140.4 |
| Episode: 3850 | Mean Train Rewards:   137.7 | Mean Test Rewards:   147.2 |
Reached reward train threshold in 3859 episodes
| Episode: 3860 | Mean Train Rewards:   152.3 | Mean Test Rewards:   139.3 |
Reached reward train threshold in 3868 episodes
| Episode: 3870 | Mean Train Rewards:   158.2 | Mean Test Rewards:   144.9 |
Reached reward test threshold in 3870 episodes
Reached reward test threshold in 3875 episodes
| Episode: 3880 | Mean Train Rewards:   157.3 | Mean Test Rewards:   150.3 |
| Episode: 3890 | Mean Train Rewards:   140.7 | Mean Test Rewards:   148.9 |
| Episode: 3900 | Mean Train Rewards:   139.3 | Mean Test Rewards:   141.5 |
| Episode: 3910 | Mean Train Rewards:   141.3 | Mean Test Rewards:   138.3 |
Reached reward train threshold in 3918 episodes
| Episode: 3920 | Mean Train Rewards:   142.6 | Mean Test Rewards:   134.6 |
| Episode: 3930 | Mean Train Rewards:   144.8 | Mean Test Rewards:   134.8 |
| Episode: 3940 | Mean Train Rewards:   130.5 | Mean Test Rewards:   136.1 |
| Episode: 3950 | Mean Train Rewards:   126.1 | Mean Test Rewards:   142.8 |
Reached reward train threshold in 3955 episodes
Reached reward test threshold in 3958 episodes
| Episode: 3960 | Mean Train Rewards:   128.8 | Mean Test Rewards:   145.0 |
Reached reward test threshold in 3961 episodes
Reached reward test threshold in 3964 episodes
Reached reward train threshold in 3965 episodes
Reached reward test threshold in 3966 episodes
| Episode: 3970 | Mean Train Rewards:   141.7 | Mean Test Rewards:   155.8 |
Reached reward test threshold in 3971 episodes
| Episode: 3980 | Mean Train Rewards:   127.6 | Mean Test Rewards:   163.3 |
Reached reward test threshold in 3984 episodes
Reached reward test threshold in 3987 episodes
| Episode: 3990 | Mean Train Rewards:   117.9 | Mean Test Rewards:   155.0 |
Reached reward test threshold in 3994 episodes
Reached reward test threshold in 3999 episodes
| Episode: 4000 | Mean Train Rewards:   123.1 | Mean Test Rewards:   146.4 |
Reached reward train threshold in 4004 episodes
Reached reward test threshold in 4004 episodes
Reached reward test threshold in 4006 episodes
Reached reward test threshold in 4007 episodes
Reached reward test threshold in 4009 episodes
```



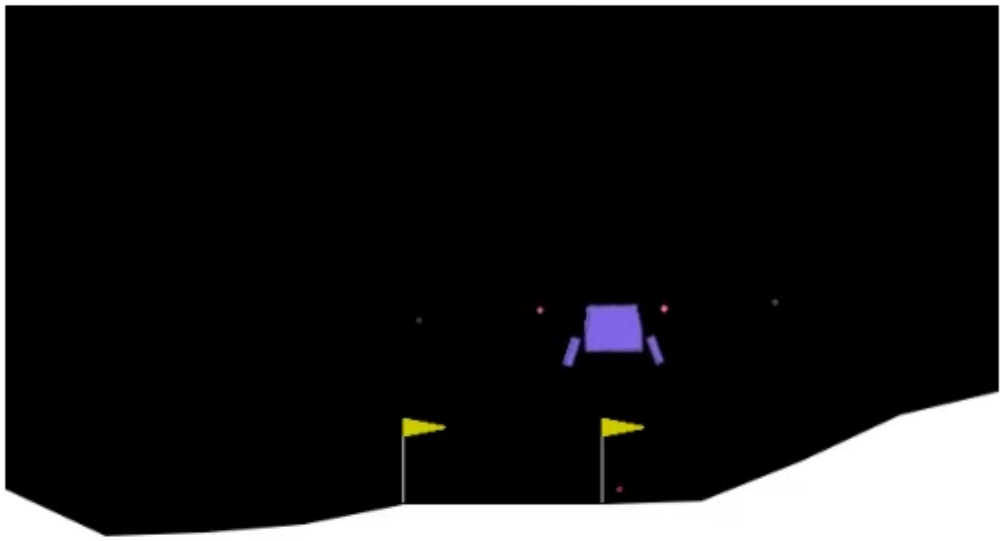
```
| Episode: 4010 | Mean Train Rewards:   132.7 | Mean Test Rewards:   167.8 |
Reached reward train threshold in 4010 episodes
Reached reward test threshold in 4010 episodes
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

```
Reached reward train threshold in 4011 episodes
Reached reward test threshold in 4011 episodes
Reached reward train threshold in 4013 episodes
Reached reward test threshold in 4013 episodes
Reached reward train threshold in 4014 episodes
Reached reward train threshold in 4018 episodes
Reached reward test threshold in 4018 episodes
Reached reward test threshold in 4019 episodes
| Episode: 4020 | Mean Train Rewards:   135.9 | Mean Test Rewards:   175.9 |
Reached reward train threshold in 4020 episodes
Reached reward test threshold in 4022 episodes
Reached reward train threshold in 4023 episodes
Reached reward test threshold in 4024 episodes
Reached reward train threshold in 4026 episodes
Reached reward train threshold in 4027 episodes
Reached reward test threshold in 4027 episodes
Reached reward train threshold in 4028 episodes
| Episode: 4030 | Mean Train Rewards:   160.9 | Mean Test Rewards:   176.9 |
Reached reward test threshold in 4032 episodes
Reached reward test threshold in 4039 episodes
| Episode: 4040 | Mean Train Rewards:   135.4 | Mean Test Rewards:   158.9 |
| Episode: 4050 | Mean Train Rewards:   120.2 | Mean Test Rewards:   145.8 |
Reached reward test threshold in 4050 episodes
Reached reward test threshold in 4053 episodes
Reached reward test threshold in 4054 episodes
Reached reward test threshold in 4056 episodes
Reached reward test threshold in 4057 episodes
Reached reward test threshold in 4058 episodes
| Episode: 4060 | Mean Train Rewards:   115.0 | Mean Test Rewards:   166.5 |
Reached reward test threshold in 4060 episodes
| Episode: 4070 | Mean Train Rewards:   131.4 | Mean Test Rewards:   164.8 |
Reached reward train threshold in 4078 episodes
| Episode: 4080 | Mean Train Rewards:   145.3 | Mean Test Rewards:   150.2 |
Reached reward train threshold in 4080 episodes
Reached reward test threshold in 4083 episodes
Reached reward train threshold in 4084 episodes
| Episode: 4090 | Mean Train Rewards:   157.3 | Mean Test Rewards:   133.2 |
Reached reward train threshold in 4093 episodes
| Episode: 4100 | Mean Train Rewards:   161.4 | Mean Test Rewards:   128.6 |
| Episode: 4110 | Mean Train Rewards:   138.5 | Mean Test Rewards:   125.8 |
| Episode: 4120 | Mean Train Rewards:   133.4 | Mean Test Rewards:   132.8 |
| Episode: 4130 | Mean Train Rewards:   137.7 | Mean Test Rewards:   138.8 |
| Episode: 4140 | Mean Train Rewards:   141.3 | Mean Test Rewards:   139.6 |
Reached reward test threshold in 4149 episodes
| Episode: 4150 | Mean Train Rewards:   139.0 | Mean Test Rewards:   132.8 |
Reached reward test threshold in 4151 episodes
Reached reward test threshold in 4153 episodes
Reached reward test threshold in 4157 episodes
| Episode: 4160 | Mean Train Rewards:   143.5 | Mean Test Rewards:   137.8 |
Reached reward test threshold in 4161 episodes
| Episode: 4170 | Mean Train Rewards:   140.1 | Mean Test Rewards:   145.8 |
| Episode: 4180 | Mean Train Rewards:   142.3 | Mean Test Rewards:   140.4 |
| Episode: 4190 | Mean Train Rewards:   142.7 | Mean Test Rewards:   132.1 |
| Episode: 4200 | Mean Train Rewards:   147.4 | Mean Test Rewards:   132.3 |
Reached reward test threshold in 4201 episodes
| Episode: 4210 | Mean Train Rewards:   146.8 | Mean Test Rewards:   140.2 |
| Episode: 4220 | Mean Train Rewards:   142.7 | Mean Test Rewards:   136.1 |
Reached reward train threshold in 4228 episodes
| Episode: 4230 | Mean Train Rewards:   153.7 | Mean Test Rewards:   127.0 |
| Episode: 4240 | Mean Train Rewards:   157.0 | Mean Test Rewards:   121.9 |
| Episode: 4250 | Mean Train Rewards:   149.7 | Mean Test Rewards:   115.8 |
| Episode: 4260 | Mean Train Rewards:   150.0 | Mean Test Rewards:   116.4 |
| Episode: 4270 | Mean Train Rewards:   147.1 | Mean Test Rewards:   130.0 |
Reached reward test threshold in 4276 episodes
| Episode: 4280 | Mean Train Rewards:   147.0 | Mean Test Rewards:   145.5 |
| Episode: 4290 | Mean Train Rewards:   142.8 | Mean Test Rewards:   143.3 |
| Episode: 4300 | Mean Train Rewards:   143.6 | Mean Test Rewards:   137.6 |
| Episode: 4310 | Mean Train Rewards:   144.3 | Mean Test Rewards:   126.4 |
| Episode: 4320 | Mean Train Rewards:   142.7 | Mean Test Rewards:   128.3 |
```

```
| Episode: 4330 | Mean Train Rewards:   138.8 | Mean Test Rewards:   126.2 |
Reached reward test threshold in 4338 episodes
Reached reward test threshold in 4339 episodes
| Episode: 4340 | Mean Train Rewards:   138.4 | Mean Test Rewards:   138.2 |
Reached reward test threshold in 4341 episodes
Reached reward test threshold in 4343 episodes
Reached reward test threshold in 4345 episodes
Reached reward test threshold in 4346 episodes
Reached reward test threshold in 4347 episodes
Reached reward test threshold in 4348 episodes
Reached reward test threshold in 4349 episodes
| Episode: 4350 | Mean Train Rewards:   136.6 | Mean Test Rewards:   175.1 |
Reached reward test threshold in 4350 episodes
Reached reward train threshold in 4352 episodes
Reached reward test threshold in 4356 episodes
| Episode: 4360 | Mean Train Rewards:   148.2 | Mean Test Rewards:   180.9 |
| Episode: 4370 | Mean Train Rewards:   149.2 | Mean Test Rewards:   162.3 |
| Episode: 4380 | Mean Train Rewards:   137.5 | Mean Test Rewards:   142.5 |
| Episode: 4390 | Mean Train Rewards:   135.8 | Mean Test Rewards:   132.4 |
| Episode: 4400 | Mean Train Rewards:   129.7 | Mean Test Rewards:   126.7 |
Reached reward train threshold in 4408 episodes
| Episode: 4410 | Mean Train Rewards:   132.9 | Mean Test Rewards:   129.6 |
Reached reward test threshold in 4411 episodes
Reached reward test threshold in 4412 episodes
Reached reward test threshold in 4413 episodes
Reached reward test threshold in 4414 episodes
Reached reward test threshold in 4415 episodes
Reached reward test threshold in 4416 episodes
Reached reward test threshold in 4417 episodes
Reached reward test threshold in 4419 episodes
| Episode: 4420 | Mean Train Rewards:   135.6 | Mean Test Rewards:   178.0 |
Reached reward test threshold in 4420 episodes
Reached reward train threshold in 4421 episodes
Reached reward test threshold in 4422 episodes
Reached reward test threshold in 4423 episodes
Reached reward train threshold in 4425 episodes
Reached reward test threshold in 4425 episodes
Reached reward train threshold in 4426 episodes
Reached reward test threshold in 4426 episodes
Reached reward train threshold in 4427 episodes
| Episode: 4430 | Mean Train Rewards:   151.6 | Mean Test Rewards:   188.0 |
Reached reward test threshold in 4430 episodes
Reached reward train threshold in 4431 episodes
Reached reward test threshold in 4432 episodes
Reached reward train threshold in 4434 episodes
Reached reward test threshold in 4435 episodes
Reached reward train threshold in 4436 episodes
Reached reward test threshold in 4436 episodes
Reached reward train threshold in 4437 episodes
Reached reward test threshold in 4437 episodes
Reached reward test threshold in 4438 episodes
| Episode: 4440 | Mean Train Rewards:   150.0 | Mean Test Rewards:   184.2 |
Reached reward train threshold in 4441 episodes
Reached reward test threshold in 4441 episodes
Reached reward train threshold in 4442 episodes
Reached reward test threshold in 4442 episodes
Reached reward train threshold in 4443 episodes
Reached reward test threshold in 4445 episodes
Reached reward test threshold in 4446 episodes
Reached reward test threshold in 4447 episodes
Reached reward train threshold in 4448 episodes
Reached reward train threshold in 4449 episodes
Reached reward test threshold in 4449 episodes
| Episode: 4450 | Mean Train Rewards:   153.2 | Mean Test Rewards:   173.0 |
Reached reward test threshold in 4451 episodes
Reached reward train threshold in 4453 episodes
Reached reward train threshold in 4454 episodes
Reached reward test threshold in 4455 episodes
Reached reward test threshold in 4456 episodes
```

```
Reached reward test threshold in 4457 episodes
Reached reward test threshold in 4458 episodes
Reached reward test threshold in 4459 episodes
| Episode: 4460 | Mean Train Rewards:   140.6 | Mean Test Rewards:   191.0 |
Reached reward test threshold in 4460 episodes
Reached reward train threshold in 4461 episodes
Reached reward test threshold in 4461 episodes
Reached reward train threshold in 4462 episodes
Reached reward test threshold in 4462 episodes
Reached reward train threshold in 4464 episodes
Reached reward test threshold in 4465 episodes
Reached reward train threshold in 4467 episodes
Reached reward test threshold in 4468 episodes
| Episode: 4470 | Mean Train Rewards:   141.1 | Mean Test Rewards:   183.4 |
Reached reward test threshold in 4470 episodes
Reached reward test threshold in 4471 episodes
Reached reward train threshold in 4472 episodes
Reached reward test threshold in 4472 episodes
Reached reward train threshold in 4473 episodes
Reached reward test threshold in 4473 episodes
Reached reward test threshold in 4474 episodes
Reached reward test threshold in 4475 episodes
Reached reward test threshold in 4476 episodes
Reached reward train threshold in 4477 episodes
Reached reward test threshold in 4477 episodes
Reached reward train threshold in 4478 episodes
Reached reward test threshold in 4478 episodes
Reached reward train threshold in 4479 episodes
Reached reward test threshold in 4479 episodes
| Episode: 4480 | Mean Train Rewards:   149.1 | Mean Test Rewards:   208.9 |
Reached reward test threshold in 4480 episodes
Reached reward test threshold in 4481 episodes
Reached reward test threshold in 4482 episodes
Reached reward test threshold in 4483 episodes
Reached reward test threshold in 4484 episodes
Reached reward test threshold in 4485 episodes
Reached reward test threshold in 4487 episodes
Reached reward test threshold in 4488 episodes
Reached reward train threshold in 4489 episodes
Reached reward test threshold in 4489 episodes
| Episode: 4490 | Mean Train Rewards:   161.8 | Mean Test Rewards:   219.2 |
Reached reward train threshold in 4490 episodes
Reached reward test threshold in 4491 episodes
Reached reward test threshold in 4492 episodes
Reached reward test threshold in 4494 episodes
Reached reward test threshold in 4496 episodes
Reached reward train threshold in 4497 episodes
Reached reward test threshold in 4497 episodes
Reached reward test threshold in 4498 episodes
Reached reward test threshold in 4499 episodes
```



```
| Episode: 4500 | Mean Train Rewards:   150.4 | Mean Test Rewards:   226.4 |
Reached reward test threshold in 4500 episodes
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarnin
g: WARN: The argument mode in render method is deprecated; use render_mode during environment
initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
```

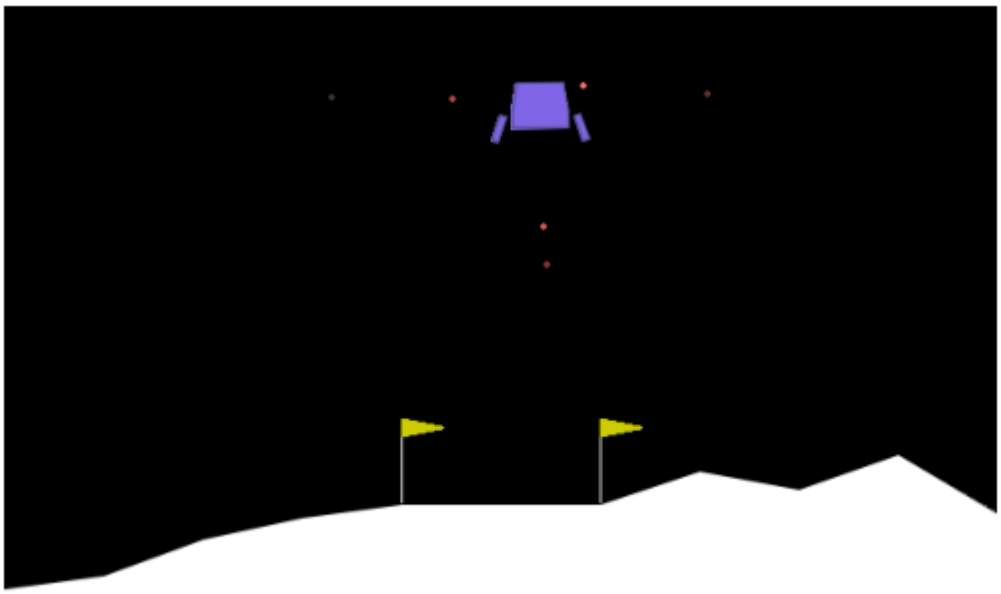
```
Reached reward test threshold in 4501 episodes
Reached reward test threshold in 4502 episodes
Reached reward test threshold in 4503 episodes
Reached reward test threshold in 4504 episodes
Reached reward test threshold in 4506 episodes
Reached reward test threshold in 4507 episodes
Reached reward test threshold in 4508 episodes
Reached reward test threshold in 4509 episodes
| Episode: 4510 | Mean Train Rewards:   145.1 | Mean Test Rewards:   218.9 |
Reached reward test threshold in 4511 episodes
Reached reward test threshold in 4512 episodes
Reached reward test threshold in 4513 episodes
Reached reward test threshold in 4514 episodes
Reached reward test threshold in 4515 episodes
Reached reward test threshold in 4516 episodes
| Episode: 4520 | Mean Train Rewards:   134.3 | Mean Test Rewards:   213.9 |
| Episode: 4530 | Mean Train Rewards:   127.2 | Mean Test Rewards:   171.5 |
| Episode: 4540 | Mean Train Rewards:   134.3 | Mean Test Rewards:   128.7 |
| Episode: 4550 | Mean Train Rewards:   142.7 | Mean Test Rewards:   135.3 |
| Episode: 4560 | Mean Train Rewards:   149.4 | Mean Test Rewards:   138.9 |
Reached reward train threshold in 4562 episodes
Reached reward train threshold in 4564 episodes
Reached reward test threshold in 4568 episodes
| Episode: 4570 | Mean Train Rewards:   164.0 | Mean Test Rewards:   141.0 |
Reached reward train threshold in 4573 episodes
| Episode: 4580 | Mean Train Rewards:   170.6 | Mean Test Rewards:   134.8 |
| Episode: 4590 | Mean Train Rewards:   152.4 | Mean Test Rewards:   132.4 |
Reached reward train threshold in 4597 episodes
| Episode: 4600 | Mean Train Rewards:   154.3 | Mean Test Rewards:   138.4 |
| Episode: 4610 | Mean Train Rewards:   156.7 | Mean Test Rewards:   142.7 |
| Episode: 4620 | Mean Train Rewards:   148.9 | Mean Test Rewards:   139.0 |
Reached reward test threshold in 4629 episodes
| Episode: 4630 | Mean Train Rewards:   139.7 | Mean Test Rewards:   141.0 |
Reached reward test threshold in 4632 episodes
Reached reward test threshold in 4633 episodes
Reached reward test threshold in 4636 episodes
Reached reward test threshold in 4637 episodes
Reached reward test threshold in 4638 episodes
Reached reward test threshold in 4639 episodes
| Episode: 4640 | Mean Train Rewards:   139.2 | Mean Test Rewards:   165.5 |
Reached reward test threshold in 4649 episodes
| Episode: 4650 | Mean Train Rewards:   143.1 | Mean Test Rewards:   166.4 |
Reached reward test threshold in 4655 episodes
Reached reward test threshold in 4657 episodes
| Episode: 4660 | Mean Train Rewards:   140.9 | Mean Test Rewards:   163.0 |
Reached reward test threshold in 4666 episodes
| Episode: 4670 | Mean Train Rewards:   134.5 | Mean Test Rewards:   159.0 |
Reached reward test threshold in 4671 episodes
| Episode: 4680 | Mean Train Rewards:   147.0 | Mean Test Rewards:   157.5 |
| Episode: 4690 | Mean Train Rewards:   143.3 | Mean Test Rewards:   146.7 |
Reached reward test threshold in 4699 episodes
| Episode: 4700 | Mean Train Rewards:   145.3 | Mean Test Rewards:   149.8 |
Reached reward test threshold in 4700 episodes
Reached reward test threshold in 4702 episodes
Reached reward test threshold in 4703 episodes
| Episode: 4710 | Mean Train Rewards:   144.3 | Mean Test Rewards:   148.4 |
Reached reward test threshold in 4715 episodes
| Episode: 4720 | Mean Train Rewards:   139.7 | Mean Test Rewards:   158.0 |
| Episode: 4730 | Mean Train Rewards:   137.3 | Mean Test Rewards:   141.6 |
| Episode: 4740 | Mean Train Rewards:   147.3 | Mean Test Rewards:   138.9 |
| Episode: 4750 | Mean Train Rewards:   155.3 | Mean Test Rewards:   131.8 |
| Episode: 4760 | Mean Train Rewards:   153.3 | Mean Test Rewards:   139.2 |
| Episode: 4770 | Mean Train Rewards:   148.1 | Mean Test Rewards:   137.9 |
| Episode: 4780 | Mean Train Rewards:   145.3 | Mean Test Rewards:   140.8 |
Reached reward test threshold in 4786 episodes
| Episode: 4790 | Mean Train Rewards:   145.4 | Mean Test Rewards:   143.3 |
Reached reward train threshold in 4797 episodes
| Episode: 4800 | Mean Train Rewards:   162.4 | Mean Test Rewards:   141.1 |
Reached reward train threshold in 4804 episodes
```

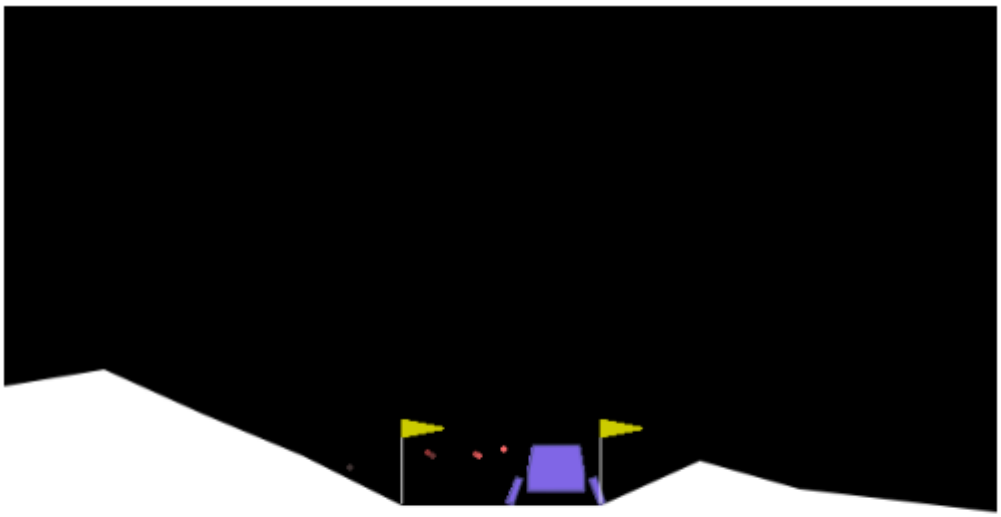
```
| Episode: 4810 | Mean Train Rewards:   172.5 | Mean Test Rewards:   144.4 |
| Episode: 4820 | Mean Train Rewards:   166.7 | Mean Test Rewards:   140.9 |
| Episode: 4830 | Mean Train Rewards:   152.1 | Mean Test Rewards:   145.5 |
Reached reward test threshold in 4831 episodes
Reached reward test threshold in 4834 episodes
Reached reward train threshold in 4835 episodes
Reached reward test threshold in 4835 episodes
Reached reward test threshold in 4836 episodes
Reached reward test threshold in 4838 episodes
Reached reward test threshold in 4839 episodes
| Episode: 4840 | Mean Train Rewards:   153.6 | Mean Test Rewards:   173.3 |
Reached reward test threshold in 4842 episodes
Reached reward test threshold in 4843 episodes
Reached reward test threshold in 4846 episodes
Reached reward test threshold in 4848 episodes
Reached reward test threshold in 4849 episodes
| Episode: 4850 | Mean Train Rewards:   158.0 | Mean Test Rewards:   206.9 |
Reached reward test threshold in 4850 episodes
Reached reward test threshold in 4851 episodes
Reached reward test threshold in 4855 episodes
Reached reward test threshold in 4856 episodes
Reached reward test threshold in 4858 episodes
Reached reward test threshold in 4859 episodes
| Episode: 4860 | Mean Train Rewards:   153.6 | Mean Test Rewards:   219.5 |
Reached reward test threshold in 4860 episodes
Reached reward test threshold in 4864 episodes
Reached reward test threshold in 4869 episodes
| Episode: 4870 | Mean Train Rewards:   154.3 | Mean Test Rewards:   213.1 |
Reached reward test threshold in 4870 episodes
Reached reward test threshold in 4871 episodes
Reached reward test threshold in 4872 episodes
Reached reward test threshold in 4873 episodes
Reached reward test threshold in 4874 episodes
Reached reward test threshold in 4876 episodes
Reached reward test threshold in 4878 episodes
| Episode: 4880 | Mean Train Rewards:   150.2 | Mean Test Rewards:   204.8 |
Reached reward test threshold in 4883 episodes
Reached reward test threshold in 4884 episodes
Reached reward test threshold in 4885 episodes
Reached reward test threshold in 4886 episodes
Reached reward test threshold in 4887 episodes
Reached reward test threshold in 4888 episodes
Reached reward test threshold in 4889 episodes
| Episode: 4890 | Mean Train Rewards:   156.0 | Mean Test Rewards:   209.8 |
Reached reward test threshold in 4890 episodes
Reached reward test threshold in 4891 episodes
Reached reward test threshold in 4892 episodes
Reached reward test threshold in 4893 episodes
Reached reward test threshold in 4894 episodes
Reached reward test threshold in 4895 episodes
Reached reward test threshold in 4896 episodes
Reached reward test threshold in 4897 episodes
Reached reward test threshold in 4898 episodes
Reached reward test threshold in 4899 episodes
| Episode: 4900 | Mean Train Rewards:   151.4 | Mean Test Rewards:   226.7 |
Reached reward test threshold in 4900 episodes
Reached reward test threshold in 4901 episodes
Reached reward test threshold in 4902 episodes
Reached reward test threshold in 4903 episodes
Reached reward test threshold in 4904 episodes
Reached reward test threshold in 4905 episodes
Reached reward test threshold in 4906 episodes
Reached reward test threshold in 4907 episodes
Reached reward test threshold in 4908 episodes
Reached reward test threshold in 4909 episodes
| Episode: 4910 | Mean Train Rewards:   152.0 | Mean Test Rewards:   255.7 |
Reached reward test threshold in 4910 episodes
Reached reward test threshold in 4911 episodes
Reached reward test threshold in 4912 episodes
```

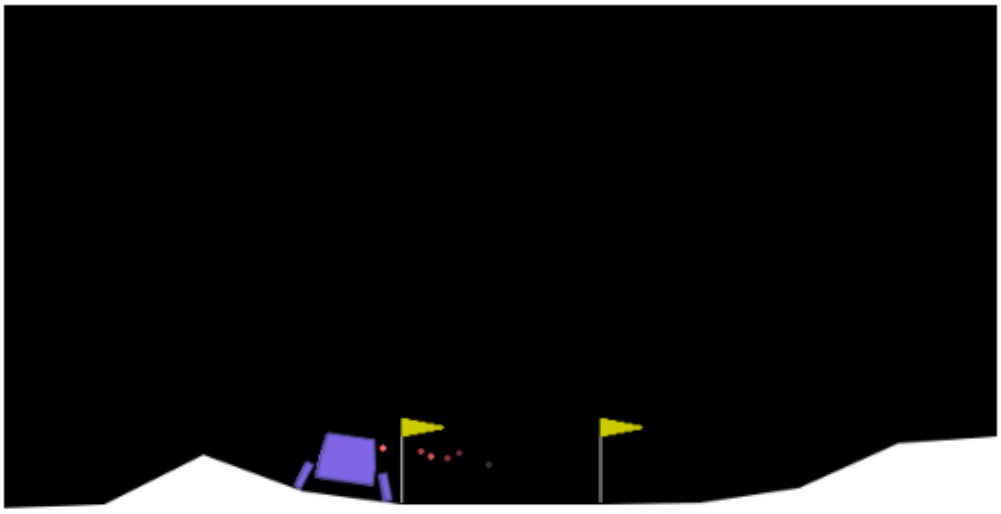
```
Reached reward test threshold in 4913 episodes
Reached reward test threshold in 4914 episodes
Reached reward test threshold in 4915 episodes
Reached reward test threshold in 4916 episodes
Reached reward test threshold in 4917 episodes
Reached reward test threshold in 4918 episodes
Reached reward test threshold in 4919 episodes
| Episode: 4920 | Mean Train Rewards:   153.9 | Mean Test Rewards:   256.5 |
Reached reward test threshold in 4920 episodes
Reached reward test threshold in 4921 episodes
Reached reward test threshold in 4922 episodes
Reached reward train threshold in 4923 episodes
Reached reward test threshold in 4923 episodes
Reached reward test threshold in 4924 episodes
Reached reward test threshold in 4925 episodes
Reached reward test threshold in 4926 episodes
Reached reward test threshold in 4927 episodes
Reached reward test threshold in 4928 episodes
Reached reward test threshold in 4929 episodes
| Episode: 4930 | Mean Train Rewards:   161.9 | Mean Test Rewards:   253.1 |
Reached reward test threshold in 4930 episodes
Reached reward test threshold in 4931 episodes
Reached reward test threshold in 4932 episodes
Reached reward test threshold in 4933 episodes
Reached reward test threshold in 4934 episodes
Reached reward test threshold in 4935 episodes
Reached reward test threshold in 4936 episodes
Reached reward test threshold in 4937 episodes
Reached reward test threshold in 4938 episodes
Reached reward test threshold in 4939 episodes
| Episode: 4940 | Mean Train Rewards:   160.7 | Mean Test Rewards:   251.5 |
Reached reward test threshold in 4940 episodes
Reached reward test threshold in 4941 episodes
Reached reward test threshold in 4942 episodes
Reached reward test threshold in 4943 episodes
Reached reward test threshold in 4944 episodes
Reached reward test threshold in 4945 episodes
Reached reward test threshold in 4946 episodes
Reached reward train threshold in 4947 episodes
Reached reward test threshold in 4947 episodes
Reached reward test threshold in 4948 episodes
Reached reward train threshold in 4949 episodes
Reached reward test threshold in 4949 episodes
| Episode: 4950 | Mean Train Rewards:   168.8 | Mean Test Rewards:   257.4 |
Reached reward train threshold in 4950 episodes
Reached reward test threshold in 4950 episodes
Reached reward train threshold in 4951 episodes
Reached reward test threshold in 4951 episodes
Reached reward test threshold in 4952 episodes
Reached reward train threshold in 4953 episodes
Reached reward test threshold in 4953 episodes
Reached reward train threshold in 4954 episodes
Reached reward test threshold in 4954 episodes
Reached reward train threshold in 4955 episodes
Reached reward test threshold in 4955 episodes
Reached reward train threshold in 4956 episodes
Reached reward test threshold in 4956 episodes
Reached reward train threshold in 4957 episodes
Reached reward test threshold in 4957 episodes
Reached reward train threshold in 4958 episodes
Reached reward test threshold in 4958 episodes
Reached reward train threshold in 4959 episodes
Reached reward test threshold in 4959 episodes
| Episode: 4960 | Mean Train Rewards:   202.2 | Mean Test Rewards:   256.6 |
Reached reward train threshold in 4960 episodes
Reached reward test threshold in 4960 episodes
Reached reward test threshold in 4961 episodes
Reached reward train threshold in 4962 episodes
Reached reward test threshold in 4962 episodes
```

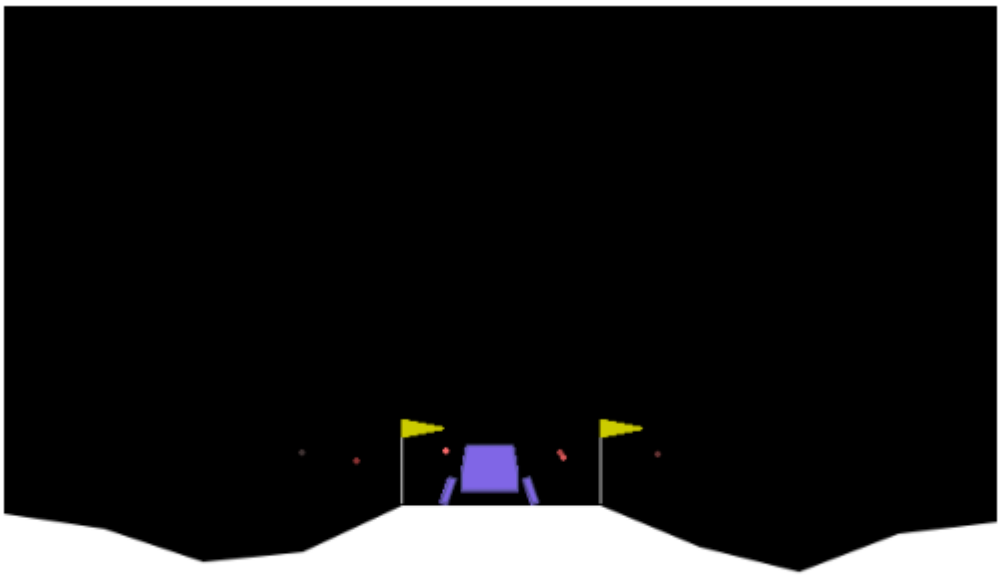
```
Reached reward train threshold in 4963 episodes
Reached reward test threshold in 4963 episodes
Reached reward train threshold in 4964 episodes
Reached reward test threshold in 4964 episodes
Reached reward train threshold in 4965 episodes
Reached reward test threshold in 4965 episodes
Reached reward train threshold in 4966 episodes
Reached reward test threshold in 4966 episodes
Reached reward train threshold in 4967 episodes
Reached reward test threshold in 4967 episodes
Reached reward train threshold in 4968 episodes
Reached reward test threshold in 4968 episodes
Reached reward train threshold in 4969 episodes
Reached reward test threshold in 4969 episodes
| Episode: 4970 | Mean Train Rewards:   244.2 | Mean Test Rewards:   255.4 |
Reached reward train threshold in 4970 episodes
Reached reward test threshold in 4970 episodes
Reached reward train threshold in 4971 episodes
Reached reward test threshold in 4971 episodes
Reached reward train threshold in 4972 episodes
Reached reward test threshold in 4972 episodes
Reached reward train threshold in 4973 episodes
Reached reward test threshold in 4973 episodes
Reached reward train threshold in 4974 episodes
Reached reward test threshold in 4974 episodes
Reached reward train threshold in 4975 episodes
Reached reward test threshold in 4975 episodes
Reached reward train threshold in 4976 episodes
Reached reward test threshold in 4976 episodes
Reached reward train threshold in 4977 episodes
Reached reward test threshold in 4977 episodes
Reached reward train threshold in 4978 episodes
Reached reward test threshold in 4978 episodes
Reached reward test threshold in 4979 episodes
| Episode: 4980 | Mean Train Rewards:   254.6 | Mean Test Rewards:   259.7 |
Reached reward train threshold in 4980 episodes
Reached reward test threshold in 4980 episodes
Reached reward train threshold in 4981 episodes
Reached reward test threshold in 4981 episodes
Reached reward train threshold in 4982 episodes
Reached reward test threshold in 4982 episodes
Reached reward train threshold in 4983 episodes
Reached reward test threshold in 4983 episodes
Reached reward train threshold in 4984 episodes
Reached reward test threshold in 4984 episodes
Reached reward train threshold in 4985 episodes
Reached reward test threshold in 4985 episodes
Reached reward test threshold in 4986 episodes
Reached reward train threshold in 4987 episodes
Reached reward test threshold in 4987 episodes
Reached reward test threshold in 4988 episodes
Reached reward train threshold in 4989 episodes
Reached reward test threshold in 4989 episodes
| Episode: 4990 | Mean Train Rewards:   241.3 | Mean Test Rewards:   256.5 |
Reached reward test threshold in 4990 episodes
Reached reward train threshold in 4991 episodes
Reached reward test threshold in 4991 episodes
Reached reward train threshold in 4992 episodes
Reached reward test threshold in 4992 episodes
Reached reward train threshold in 4993 episodes
Reached reward test threshold in 4993 episodes
Reached reward train threshold in 4994 episodes
Reached reward test threshold in 4994 episodes
Reached reward train threshold in 4995 episodes
Reached reward test threshold in 4995 episodes
Reached reward test threshold in 4996 episodes
Reached reward train threshold in 4997 episodes
Reached reward test threshold in 4997 episodes
Reached reward train threshold in 4998 episodes
```

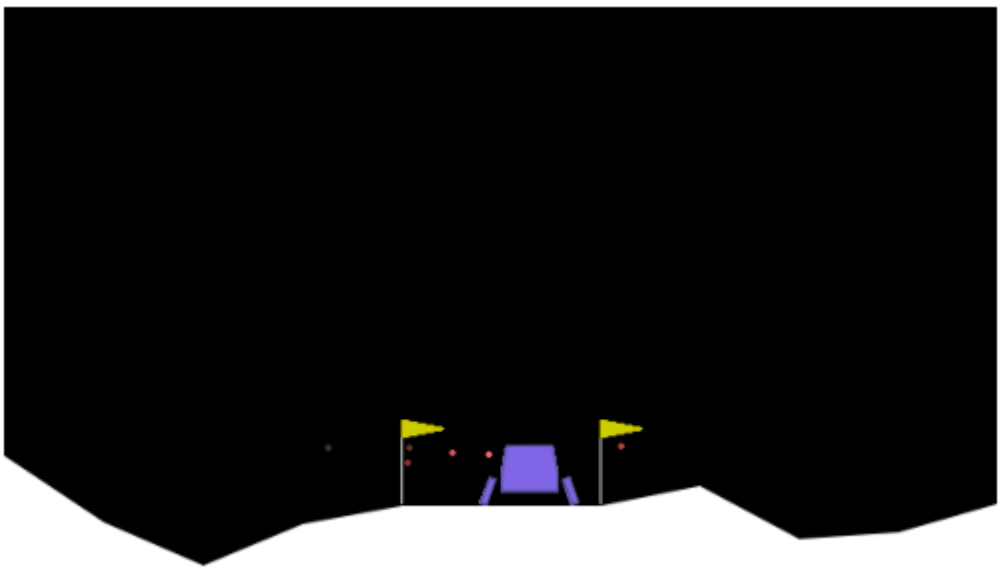
```
Reached reward test threshold in 4998 episodes
Reached reward train threshold in 4999 episodes
Reached reward test threshold in 4999 episodes
| Episode: 5000 | Mean Train Rewards:   231.2 | Mean Test Rewards:   254.4 |
Reached reward train threshold in 5000 episodes
Reached reward test threshold in 5000 episodes
```

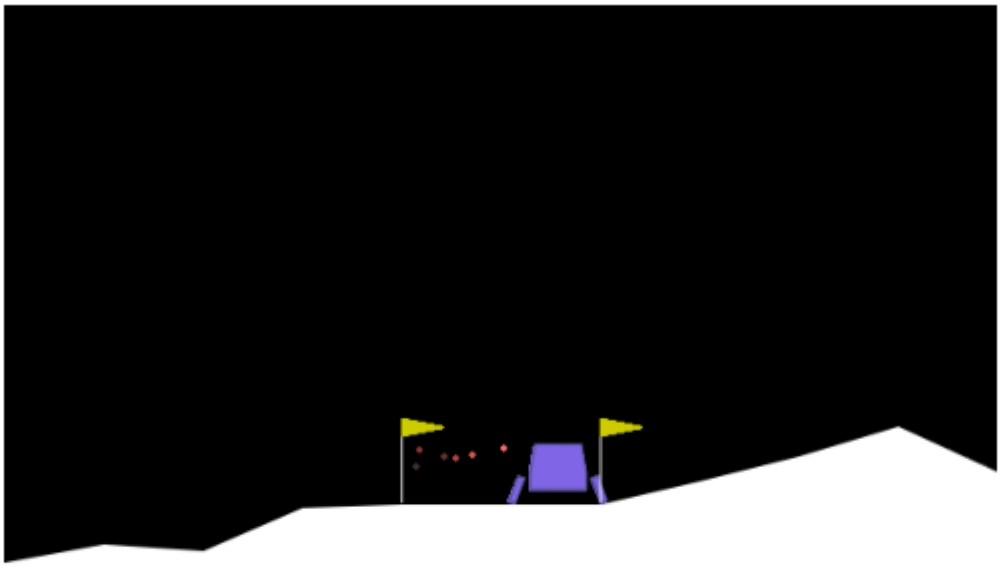


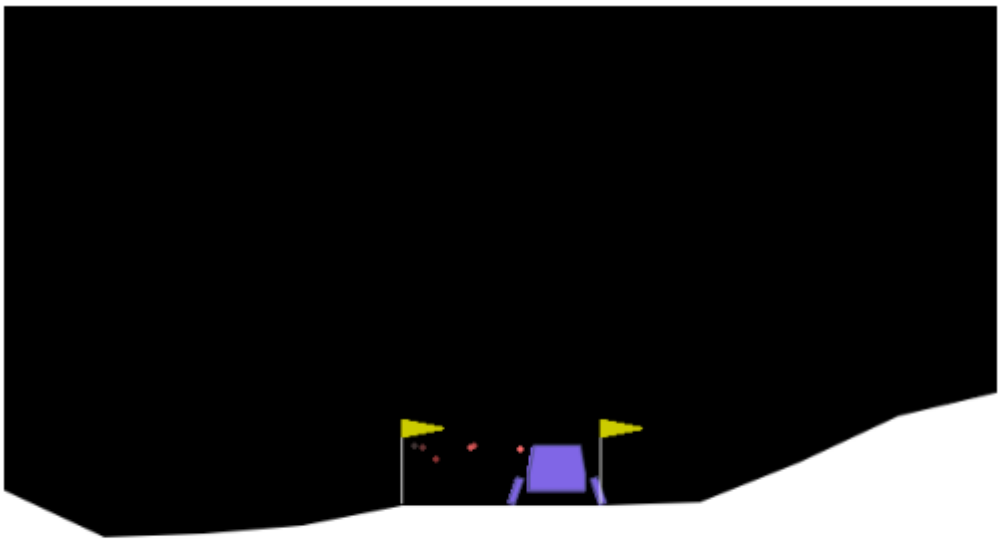








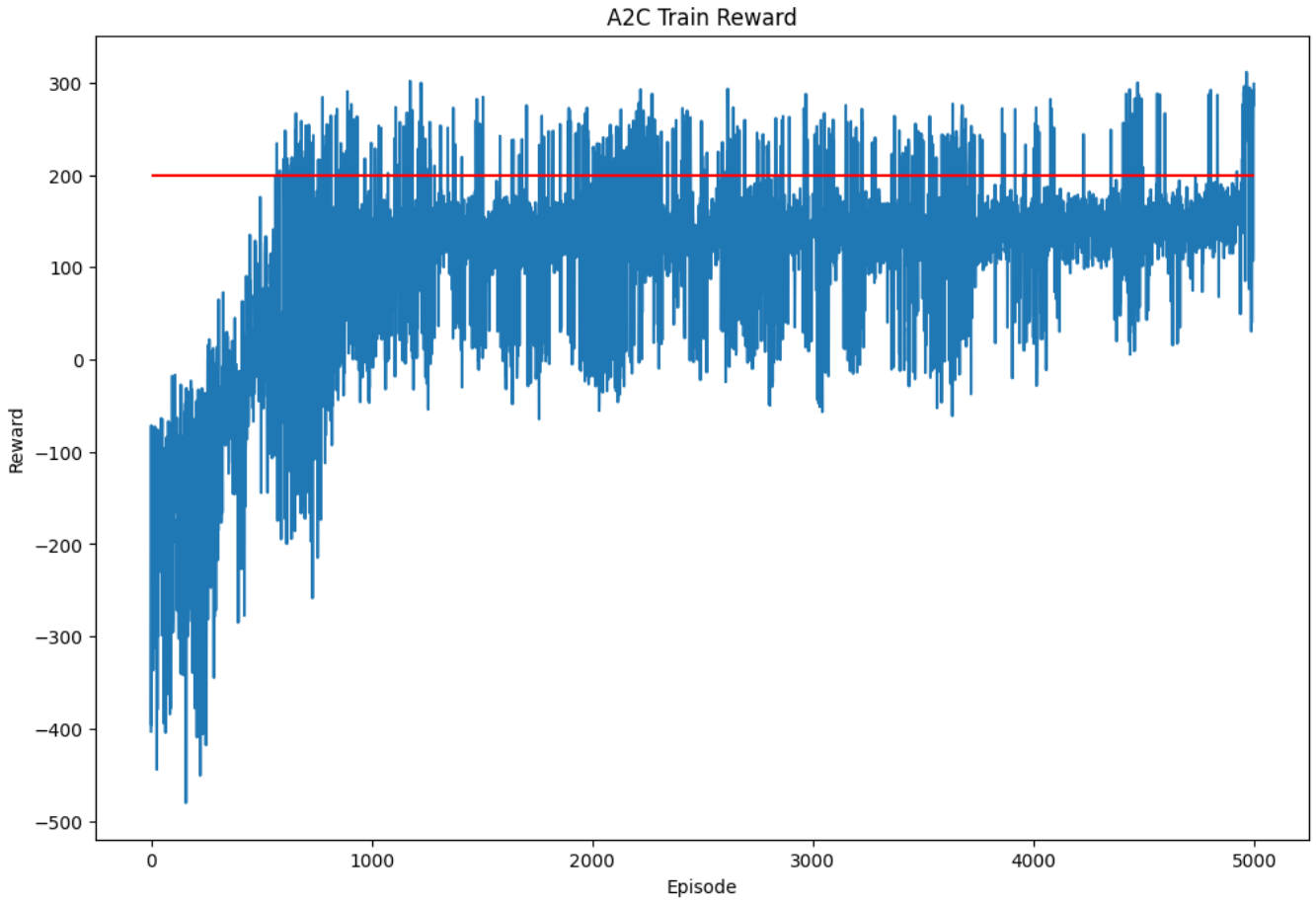




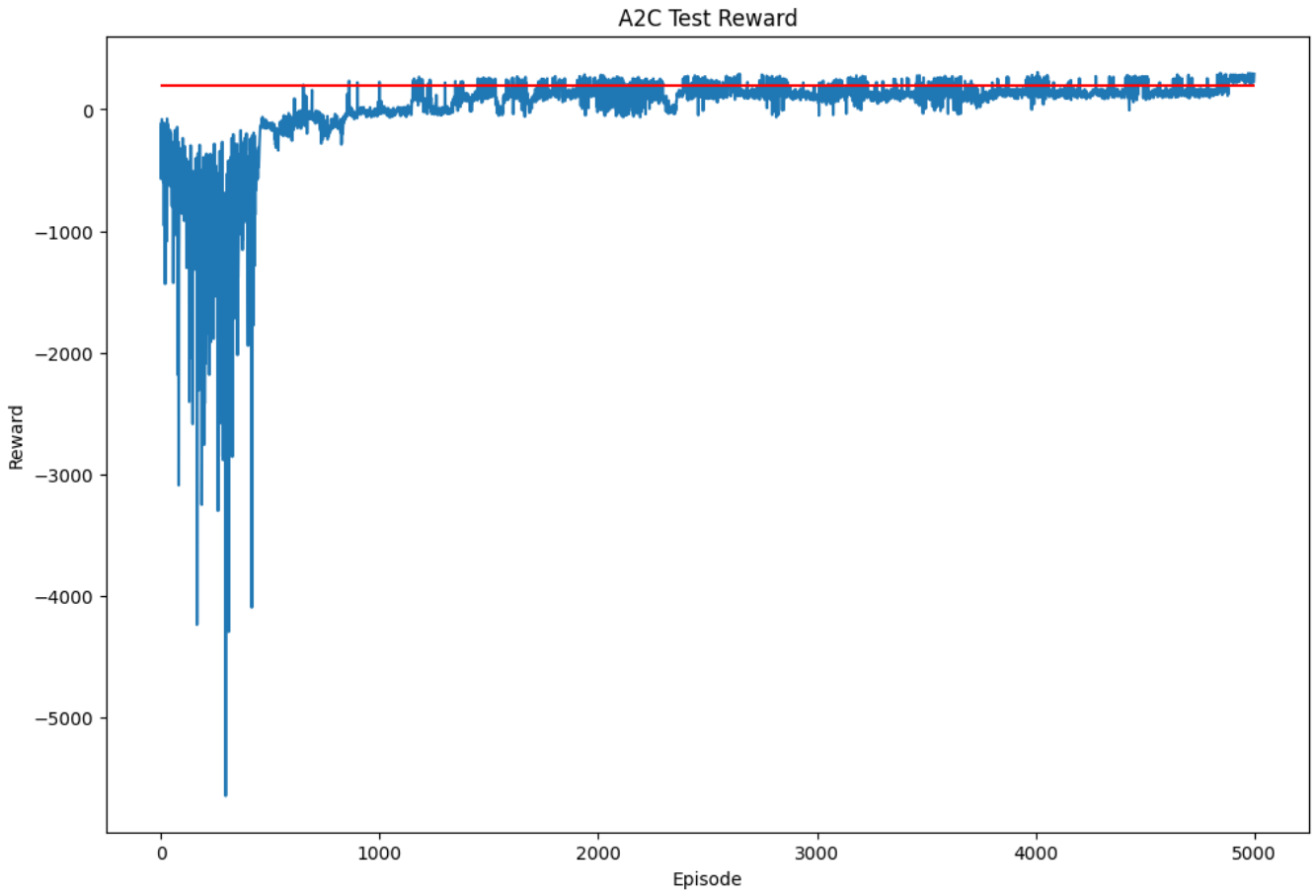
Evaluation of A2C Model

In [ ]:
```
a2c_train_rewards = np.array(a2c_train_rewards)
a2c_test_rewards = np.array(a2c_test_rewards)
```

In [ ]:
```
plt.figure(figsize=(12,8))
plt.plot(a2c_train_rewards, label='Test Reward')
plt.xlabel('Episode')
plt.ylabel('Reward')
plt.hlines(REWARD_THRESHOLD, 0, len(a2c_train_rewards), color='r')
plt.title("A2C Train Reward")
plt.show()
```



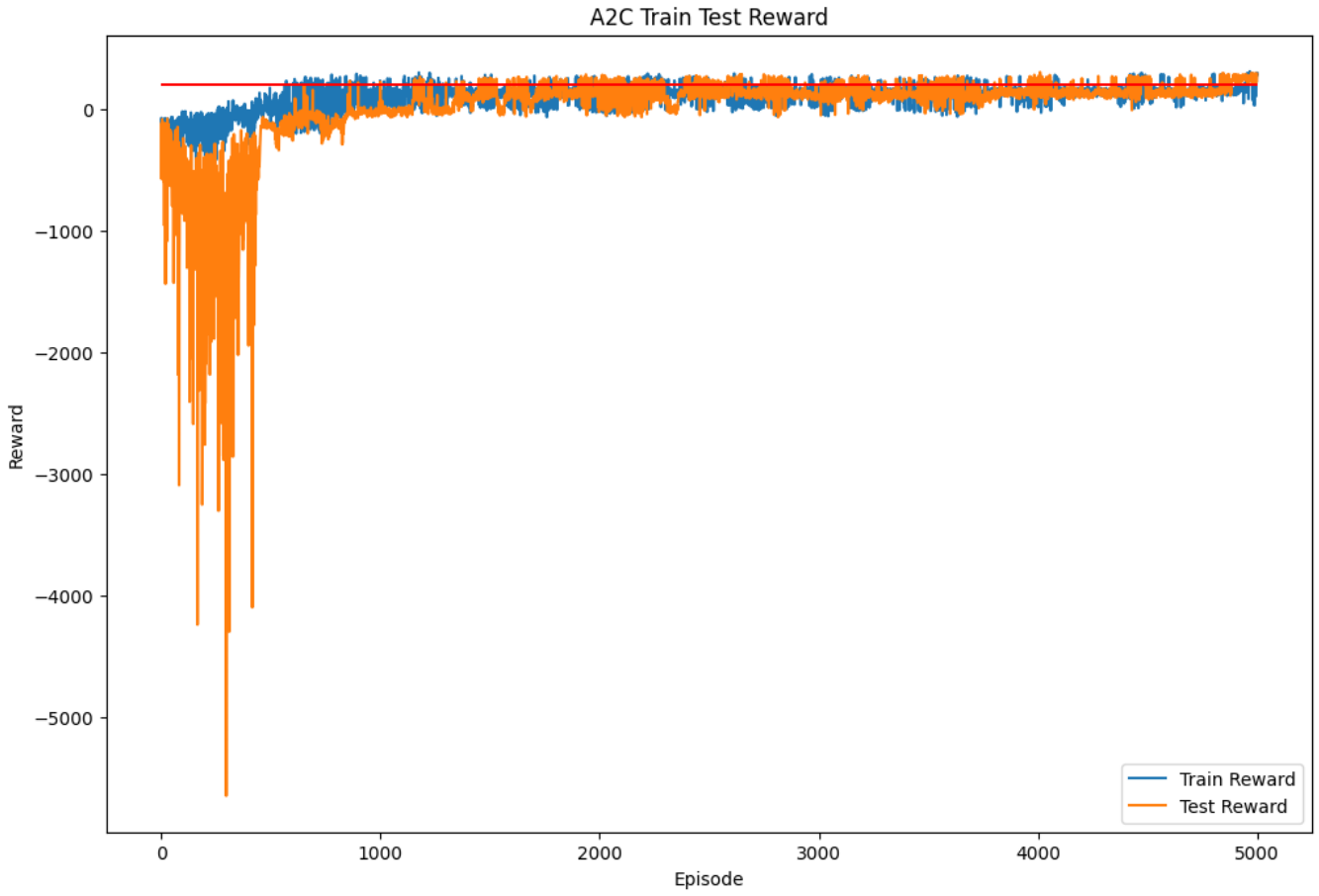
In [ ]:
```
plt.figure(figsize=(12,8))
plt.plot(a2c_test_rewards, label='Test Reward')
plt.xlabel('Episode')
plt.ylabel('Reward')
plt.hlines(REWARD_THRESHOLD, 0, len(a2c_test_rewards), color='r')
plt.title("A2C Test Reward")
plt.show()
```



In [ ]:
```
plt.figure(figsize=(12,8))
plt.plot(a2c_train_rewards, label='Train Reward')
```

```
plt.plot(a2c_test_rewards, label='Test Reward')
plt.xlabel('Episode')
plt.ylabel('Reward')
plt.hlines(REWARD_THRESHOLD, 0, len(a2c_test_rewards), color='r')
plt.title("A2C Train Test Reward")
plt.legend(loc='lower right')
```

Out[ ]:
```
<matplotlib.legend.Legend at 0x2b31c0a9fa0>
```



In [ ]:
```
print(f"Highest Train Episode: {np.argmax(a2c_train_rewards) + 1} Score: {np.max(a2c_train_re
print(f"Highest Test Episode: {np.argmax(a2c_test_rewards) + 1} Score: {np.max(a2c_test_rewar
```

```
Highest Train Episode: 4967 Score: 311.61627310306244
Highest Test Episode: 4010 Score: 304.8280801670431
```

Observation

Visualizing model

We will be choosing the values where both training and testing rewards are the higher than the reward threshold. Train >= 285 and Test >= 285

In [ ]:
```
best_episodes = np.intersect1d(np.where(
    np.array(a2c_train_rewards) >= 285), np.where(np.array(a2c_test_rewards) >= 285)) + 1
best_episodes
```

Out[ ]:
```
array([5000], dtype=int64)
```

In [ ]:
```
best_episode = best_episodes[0]
a2c_policy.load_state_dict(torch.load(
    f"./checkpoints/A2C/A2CTrain-{best_episode}.pth"))
a2c_policy
```

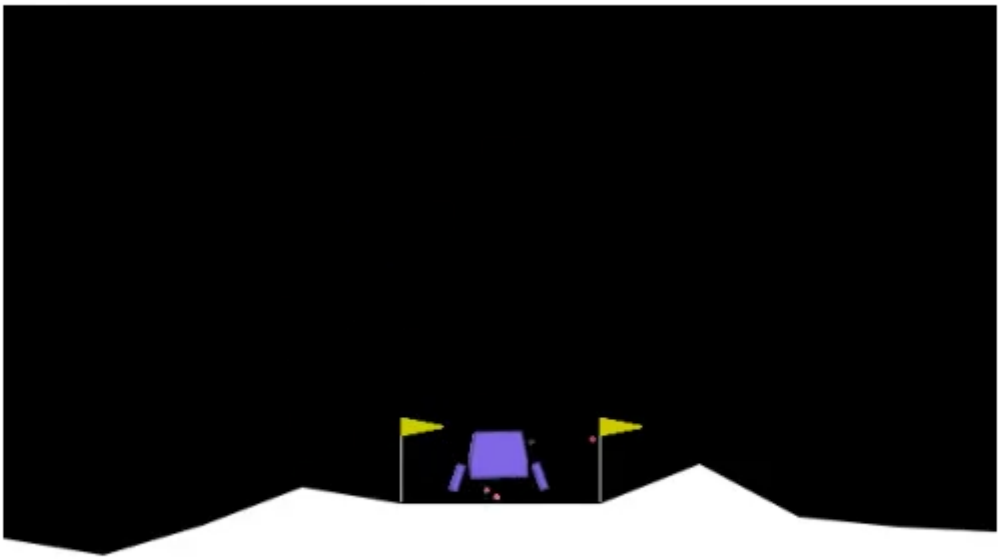
Out[ ]:
```
ActorCritic(
  (actor): MLP(
    (net): Sequential(
      (0): Linear(in_features=8, out_features=128, bias=True)
      (1): Dropout(p=0.1, inplace=False)
      (2): PReLU(num_parameters=1)
      (3): Linear(in_features=128, out_features=128, bias=True)
      (4): Dropout(p=0.1, inplace=False)
      (5): PReLU(num_parameters=1)
      (6): Linear(in_features=128, out_features=4, bias=True)
    )
  )
  (critic): MLP(
    (net): Sequential(
      (0): Linear(in_features=8, out_features=128, bias=True)
      (1): Dropout(p=0.1, inplace=False)
      (2): PReLU(num_parameters=1)
      (3): Linear(in_features=128, out_features=128, bias=True)
      (4): Dropout(p=0.1, inplace=False)
      (5): PReLU(num_parameters=1)
      (6): Linear(in_features=128, out_features=1, bias=True)
    )
  )
)
```

In [ ]:
```python
def show_a2c(env, policy, weights_path, iteration=1):
    policy.load_state_dict(torch.load(weights_path))
    policy.eval()

    rewards = []
    done = False
    episode_reward = 0
    frames = []
    state = env.reset()
    while not done:

        state = torch.FloatTensor(state).unsqueeze(0)

        with torch.no_grad():

            action_pred, _ = policy(state)

            action_prob = F.softmax(action_pred, dim=-1)

        action = torch.argmax(action_prob, dim=-1)

        state, reward, done, _ = env.step(action.item())
        screen = env.render(mode='rgb_array')
        episode_reward += reward
        frames.append(screen)
    return episode_reward, frames
```

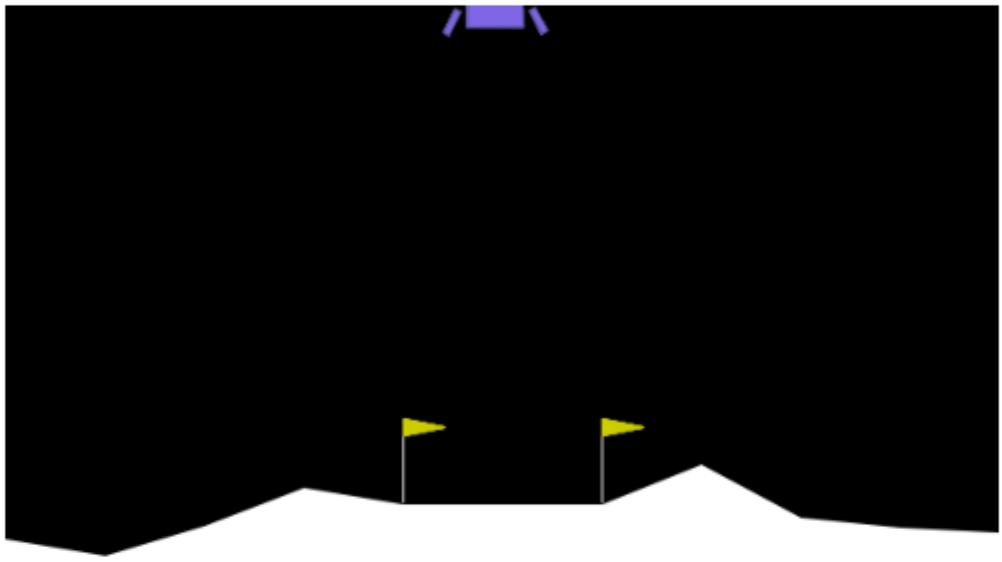
In [ ]:
```python
show_env = gym.make('LunarLander-v2',
                    continuous=False,
                    gravity=-10.0,
                    enable_wind=False,
                    wind_power=15.0,
                    turbulence_power=1.5
                    )
show_env.seed(75)
reward, frames = show_a2c(
    show_env, a2c_policy,  f"./checkpoints/A2C/A2CTrain-{best_episode}.pth")

print(f"Model Score: {reward}")

create_animation(frames, f"./videos/A2CTrain-{best_episode}-{reward}.mp4")
```

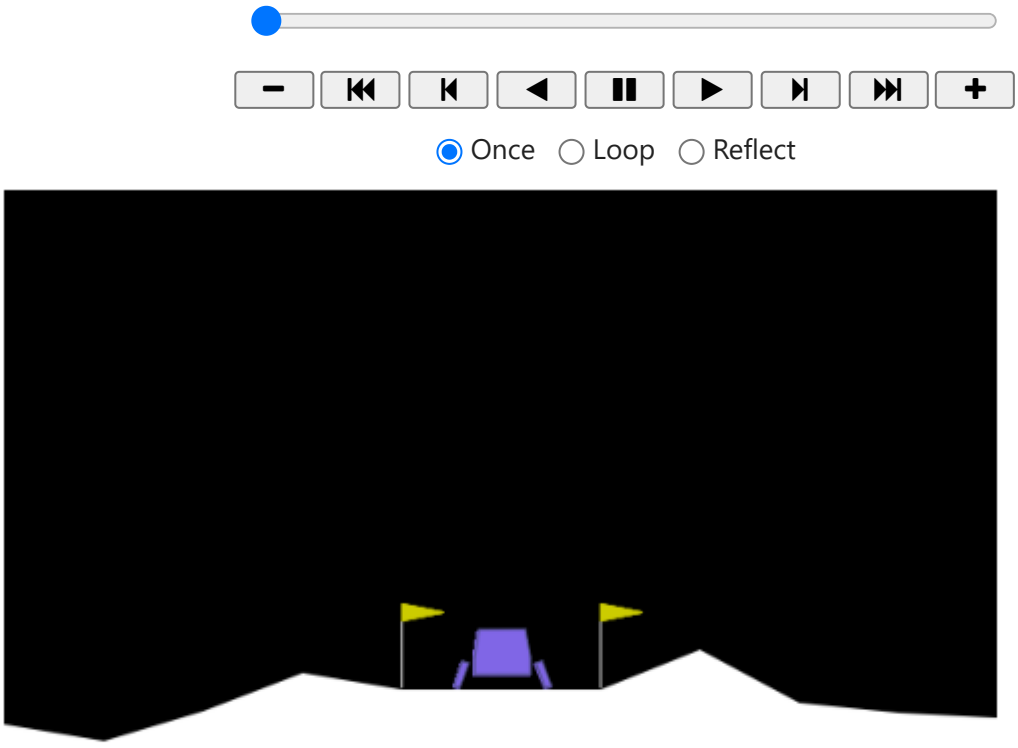
```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\utils\passive_env_checker.py:241: DeprecationWarning: `np.bool8` is a deprecated alias for `np.bool_`.  (Deprecated NumPy 1.24)
  if not isinstance(terminated, (bool, np.bool8)):
```

Model Score: 307.0810140291526



Out[ ]:





Proximal Policy Optimization

PPO, or Proximal Policy Optimization, is a type of reinforcement learning algorithm that falls under the policy gradient method category. It is an on-policy algorithm that updates the policy by optimizing a surrogate objective that is similar to the expected reward. The surrogate objective helps maintain stability during training by limiting the amount of change in the policy with each update. PPO uses a value function to compute the advantage, which is the discrepancy between the expected return of an action and the expected return of the current policy. The policy is updated by maximizing the surrogate

objective with regards to the policy network's parameters. PPO also employs a trust region optimization method that restricts the size of the policy update to further enhance stability. PPO has been extensively utilized in deep reinforcement learning and has demonstrated good performance in various scenarios.

Create environment for the PPO model

In [ ]:
```
train_env = gym.make('LunarLander-v2',
                     continuous=False,
                     gravity=-10.0,
                     enable_wind=False,
                     wind_power=15.0,
                     turbulence_power=1.5
                     )
train_env.seed(0)

test_env = gym.make('LunarLander-v2',
                    continuous=False,
                    gravity=-10.0,
                    enable_wind=False,
                    wind_power=15.0,
                    turbulence_power=1.5
                    )
test_env.seed(1)
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:317: DeprecationWarning: WARN: Initializing wrapper in old step API which returns one bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\wrappers\step_api_compatibility.py:39: DeprecationWarning: WARN: Initializing environment in old step API which returns one bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:256: DeprecationWarning: WARN: Function `env.seed(seed)` is marked as deprecated and will be removed in the future. Please use `env.reset(seed=seed)` instead.
  deprecation(
```

Out[ ]:
```
[1]
```

Setup model architecture

In [ ]:
```
def update_policy(policy, states, actions, log_prob_actions, advantages, returns, optimizer,
    total_policy_loss = 0
    total_value_loss = 0
    states = states.detach()
    actions = actions.detach()
    log_prob_actions = log_prob_actions.detach()
    advantages = advantages.detach()
    returns = returns.detach()
    for _ in range(ppo_steps):
        #get new log prob of actions for all input states
        action_pred, value_pred = policy(states)
        value_pred = value_pred.squeeze(-1)
        action_prob = F.softmax(action_pred, dim = -1)
        dist = distributions.Categorical(action_prob)
        #new log prob using old actions
        new_log_prob_actions = dist.log_prob(actions)
        policy_ratio = (new_log_prob_actions - log_prob_actions).exp()
        policy_loss_1 = policy_ratio * advantages
        policy_loss_2 = torch.clamp(policy_ratio, min = 1.0 - ppo_clip, max = 1.0 + ppo_clip)
        policy_loss = - torch.min(policy_loss_1, policy_loss_2).mean()
        value_loss = F.smooth_l1_loss(returns, value_pred).mean()
        optimizer.zero_grad()
        policy_loss.backward()
```

```
        value_loss.backward()
        optimizer.step()
        total_policy_loss += policy_loss.item()
        total_value_loss += value_loss.item()
    return total_policy_loss / ppo_steps, total_value_loss / ppo_steps
```

Training PPO Model

In [ ]:
```
def ppo_train(env, policy, optimizer, discount_factor, ppo_steps, ppo_clip):
    policy.train()
    states = []
    actions = []
    log_prob_actions = []
    values = []
    rewards = []
    done = False
    episode_reward = 0
    state = env.reset()
    while not done:
        state = torch.FloatTensor(state).unsqueeze(0)
        #append state here, not after we get the next state from env.step()
        states.append(state)
        action_pred, value_pred = policy(state)
        action_prob = F.softmax(action_pred, dim = -1)
        dist = distributions.Categorical(action_prob)
        action = dist.sample()
        log_prob_action = dist.log_prob(action)
        state, reward, done, _ = env.step(action.item())
        actions.append(action)
        log_prob_actions.append(log_prob_action)
        values.append(value_pred)
        rewards.append(reward)
        episode_reward += reward
    states = torch.cat(states)
    actions = torch.cat(actions)
    log_prob_actions = torch.cat(log_prob_actions)
    values = torch.cat(values).squeeze(-1)
    returns = calculate_returns(rewards, discount_factor)
    advantages = calculate_advantages(returns, values)
    policy_loss, value_loss = update_policy(policy, states, actions, log_prob_actions, advant
    return policy_loss, value_loss, episode_reward
```

In [ ]:
```
INPUT_DIM = train_env.observation_space.shape[0]
HIDDEN_DIM = 128
OUTPUT_DIM = train_env.action_space.n

actor = MLP(INPUT_DIM, HIDDEN_DIM, OUTPUT_DIM)
critic = MLP(INPUT_DIM, HIDDEN_DIM, 1)

ppo_policy = ActorCritic(actor, critic)
```

In [ ]:
```
LEARNING_RATE = 0.0005

optimizer = optim.Adam(ppo_policy.parameters(), lr=LEARNING_RATE)
```

In [ ]:
```
def ppo_evaluate(env, policy):
    policy.eval()
    rewards = []
    done = False
    episode_reward = 0
    frames = []
    state = env.reset()
    while not done:
        state = torch.FloatTensor(state).unsqueeze(0)
        with torch.no_grad():
            action_pred, _ = policy(state)
            action_prob = F.softmax(action_pred, dim=-1)
```

```
        action = torch.argmax(action_prob, dim=-1)
        state, reward, done, _ = env.step(action.item())
        screen = env.render(mode='rgb_array')
        episode_reward += reward
        frames.append(screen)
    return episode_reward, frames
```

In [ ]:

```
MAX_EPISODES = 3_000
DISCOUNT_FACTOR = 0.99
N_TRIALS = 25
REWARD_THRESHOLD = 200
PRINT_EVERY = 10
PPO_STEPS = 5
PPO_CLIP = 0.2

ppo_train_rewards = []
ppo_test_rewards = []

for episode in range(1, MAX_EPISODES+1):

    policy_loss, value_loss, train_reward = ppo_train(
        train_env, ppo_policy, optimizer, DISCOUNT_FACTOR, PPO_STEPS, PPO_CLIP)

    test_reward, _ = ppo_evaluate(test_env, ppo_policy)

    ppo_train_rewards.append(train_reward)
    ppo_test_rewards.append(test_reward)

    mean_ppo_train_rewards = np.mean(ppo_train_rewards[-N_TRIALS:])
    mean_ppo_test_rewards = np.mean(ppo_test_rewards[-N_TRIALS:])

    if episode % PRINT_EVERY == 0:
        print(f'| Episode: {episode:3} | Mean Train Rewards: {mean_ppo_train_rewards:7.1f} |

    if train_reward >= REWARD_THRESHOLD:
        print(f'Reached reward train threshold in {episode} episodes')
        torch.save(ppo_policy.state_dict(),
                   f"./checkpoints/PPO/PPOTrain-{episode}.pth")

    if test_reward >= REWARD_THRESHOLD:
        print(f'Reached reward test threshold in {episode} episodes')
        torch.save(ppo_policy.state_dict(),
                   f"./checkpoints/PPO/PPOTest-{episode}.pth")
```

```
| Episode:  10 | Mean Train Rewards:   -60.0 | Mean Test Rewards:  -379.0 |
| Episode:  20 | Mean Train Rewards:   -15.5 | Mean Test Rewards:  -330.9 |
| Episode:  30 | Mean Train Rewards:    17.4 | Mean Test Rewards:  -258.1 |
| Episode:  40 | Mean Train Rewards:    52.5 | Mean Test Rewards:  -223.9 |
| Episode:  50 | Mean Train Rewards:    54.3 | Mean Test Rewards:  -197.1 |
| Episode:  60 | Mean Train Rewards:    14.3 | Mean Test Rewards:  -212.8 |
| Episode:  70 | Mean Train Rewards:     7.4 | Mean Test Rewards:  -162.9 |
| Episode:  80 | Mean Train Rewards:    50.6 | Mean Test Rewards:  -153.0 |
| Episode:  90 | Mean Train Rewards:    68.2 | Mean Test Rewards:  -140.7 |
| Episode: 100 | Mean Train Rewards:    86.8 | Mean Test Rewards:   -80.5 |
| Episode: 110 | Mean Train Rewards:   107.9 | Mean Test Rewards:   -20.8 |
| Episode: 120 | Mean Train Rewards:    95.1 | Mean Test Rewards:   -34.4 |
| Episode: 130 | Mean Train Rewards:    79.5 | Mean Test Rewards:   -73.8 |
| Episode: 140 | Mean Train Rewards:    68.2 | Mean Test Rewards:   -95.2 |
| Episode: 150 | Mean Train Rewards:    80.7 | Mean Test Rewards:   -93.7 |
| Episode: 160 | Mean Train Rewards:    82.7 | Mean Test Rewards:   -87.1 |
| Episode: 170 | Mean Train Rewards:    87.7 | Mean Test Rewards:   -36.3 |
| Episode: 180 | Mean Train Rewards:    95.9 | Mean Test Rewards:    -6.8 |
| Episode: 190 | Mean Train Rewards:   110.3 | Mean Test Rewards:    10.0 |
| Episode: 200 | Mean Train Rewards:   100.9 | Mean Test Rewards:     6.2 |
| Episode: 210 | Mean Train Rewards:    90.9 | Mean Test Rewards:   -27.9 |
| Episode: 220 | Mean Train Rewards:    71.3 | Mean Test Rewards:   -39.2 |
Reached reward train threshold in 228 episodes
| Episode: 230 | Mean Train Rewards:    89.9 | Mean Test Rewards:   -35.0 |
| Episode: 240 | Mean Train Rewards:    89.3 | Mean Test Rewards:   -29.2 |
| Episode: 250 | Mean Train Rewards:    96.7 | Mean Test Rewards:   -35.2 |
| Episode: 260 | Mean Train Rewards:   103.6 | Mean Test Rewards:     8.5 |
| Episode: 270 | Mean Train Rewards:   114.4 | Mean Test Rewards:    77.9 |
| Episode: 280 | Mean Train Rewards:   110.7 | Mean Test Rewards:   107.5 |
Reached reward test threshold in 281 episodes
Reached reward test threshold in 285 episodes
| Episode: 290 | Mean Train Rewards:   117.4 | Mean Test Rewards:   110.6 |
| Episode: 300 | Mean Train Rewards:   117.5 | Mean Test Rewards:    92.0 |
| Episode: 310 | Mean Train Rewards:    95.5 | Mean Test Rewards:    29.6 |
| Episode: 320 | Mean Train Rewards:    80.9 | Mean Test Rewards:     4.1 |
| Episode: 330 | Mean Train Rewards:    86.2 | Mean Test Rewards:    14.3 |
| Episode: 340 | Mean Train Rewards:    81.5 | Mean Test Rewards:    17.5 |
| Episode: 350 | Mean Train Rewards:    82.8 | Mean Test Rewards:     4.0 |
| Episode: 360 | Mean Train Rewards:    90.3 | Mean Test Rewards:   -18.7 |
| Episode: 370 | Mean Train Rewards:    85.9 | Mean Test Rewards:   -13.4 |
| Episode: 380 | Mean Train Rewards:    85.3 | Mean Test Rewards:     6.1 |
| Episode: 390 | Mean Train Rewards:    99.2 | Mean Test Rewards:   -11.2 |
| Episode: 400 | Mean Train Rewards:    67.7 | Mean Test Rewards:   -56.9 |
| Episode: 410 | Mean Train Rewards:    41.8 | Mean Test Rewards:   -51.6 |
| Episode: 420 | Mean Train Rewards:    46.1 | Mean Test Rewards:     3.7 |
| Episode: 430 | Mean Train Rewards:    86.2 | Mean Test Rewards:     7.5 |
Reached reward train threshold in 430 episodes
| Episode: 440 | Mean Train Rewards:    74.9 | Mean Test Rewards:   -18.1 |
| Episode: 450 | Mean Train Rewards:    71.7 | Mean Test Rewards:     4.0 |
Reached reward test threshold in 457 episodes
| Episode: 460 | Mean Train Rewards:    94.8 | Mean Test Rewards:    69.1 |
Reached reward test threshold in 462 episodes
| Episode: 470 | Mean Train Rewards:   111.2 | Mean Test Rewards:   108.5 |
| Episode: 480 | Mean Train Rewards:   112.0 | Mean Test Rewards:   120.2 |
Reached reward test threshold in 489 episodes
| Episode: 490 | Mean Train Rewards:   116.5 | Mean Test Rewards:    98.5 |
| Episode: 500 | Mean Train Rewards:   120.2 | Mean Test Rewards:    90.4 |
| Episode: 510 | Mean Train Rewards:   122.2 | Mean Test Rewards:    95.1 |
Reached reward test threshold in 515 episodes
| Episode: 520 | Mean Train Rewards:   123.9 | Mean Test Rewards:    93.9 |
Reached reward train threshold in 529 episodes
| Episode: 530 | Mean Train Rewards:   119.7 | Mean Test Rewards:   102.4 |
Reached reward train threshold in 535 episodes
Reached reward test threshold in 536 episodes
Reached reward test threshold in 537 episodes
| Episode: 540 | Mean Train Rewards:   116.0 | Mean Test Rewards:   101.8 |
| Episode: 550 | Mean Train Rewards:   121.1 | Mean Test Rewards:    97.4 |
Reached reward test threshold in 550 episodes
| Episode: 560 | Mean Train Rewards:   119.3 | Mean Test Rewards:    97.0 |
```

```
Reached reward test threshold in 560 episodes
| Episode: 570 | Mean Train Rewards:   113.2 | Mean Test Rewards:    88.7 |
| Episode: 580 | Mean Train Rewards:   110.8 | Mean Test Rewards:   102.2 |
| Episode: 590 | Mean Train Rewards:   111.2 | Mean Test Rewards:    90.5 |
Reached reward test threshold in 592 episodes
Reached reward train threshold in 595 episodes
| Episode: 600 | Mean Train Rewards:   110.4 | Mean Test Rewards:    93.1 |
Reached reward test threshold in 603 episodes
Reached reward test threshold in 606 episodes
Reached reward test threshold in 608 episodes
| Episode: 610 | Mean Train Rewards:   107.5 | Mean Test Rewards:    98.8 |
Reached reward train threshold in 619 episodes
| Episode: 620 | Mean Train Rewards:   109.8 | Mean Test Rewards:    82.3 |
| Episode: 630 | Mean Train Rewards:   118.5 | Mean Test Rewards:    88.5 |
| Episode: 640 | Mean Train Rewards:   118.0 | Mean Test Rewards:    79.3 |
| Episode: 650 | Mean Train Rewards:   107.2 | Mean Test Rewards:    84.7 |
| Episode: 660 | Mean Train Rewards:    96.4 | Mean Test Rewards:    74.2 |
| Episode: 670 | Mean Train Rewards:    89.8 | Mean Test Rewards:    57.2 |
Reached reward test threshold in 679 episodes
| Episode: 680 | Mean Train Rewards:    93.6 | Mean Test Rewards:    72.4 |
| Episode: 690 | Mean Train Rewards:   101.4 | Mean Test Rewards:    89.3 |
Reached reward test threshold in 690 episodes
Reached reward test threshold in 691 episodes
| Episode: 700 | Mean Train Rewards:   112.6 | Mean Test Rewards:   121.2 |
Reached reward test threshold in 700 episodes
Reached reward test threshold in 701 episodes
Reached reward test threshold in 702 episodes
Reached reward test threshold in 704 episodes
Reached reward test threshold in 705 episodes
| Episode: 710 | Mean Train Rewards:   117.4 | Mean Test Rewards:   163.3 |
Reached reward test threshold in 712 episodes
Reached reward test threshold in 713 episodes
Reached reward test threshold in 715 episodes
Reached reward test threshold in 716 episodes
| Episode: 720 | Mean Train Rewards:   118.8 | Mean Test Rewards:   194.3 |
Reached reward test threshold in 720 episodes
Reached reward test threshold in 721 episodes
Reached reward test threshold in 724 episodes
Reached reward test threshold in 725 episodes
Reached reward test threshold in 727 episodes
Reached reward test threshold in 728 episodes
Reached reward test threshold in 729 episodes
| Episode: 730 | Mean Train Rewards:   122.7 | Mean Test Rewards:   200.3 |
Reached reward train threshold in 730 episodes
Reached reward test threshold in 730 episodes
Reached reward train threshold in 731 episodes
Reached reward test threshold in 731 episodes
Reached reward train threshold in 733 episodes
Reached reward train threshold in 734 episodes
Reached reward test threshold in 734 episodes
Reached reward train threshold in 735 episodes
Reached reward test threshold in 735 episodes
Reached reward train threshold in 736 episodes
Reached reward test threshold in 737 episodes
Reached reward train threshold in 738 episodes
Reached reward test threshold in 738 episodes
Reached reward train threshold in 739 episodes
Reached reward test threshold in 739 episodes
| Episode: 740 | Mean Train Rewards:   156.6 | Mean Test Rewards:   208.9 |
Reached reward test threshold in 740 episodes
Reached reward test threshold in 741 episodes
Reached reward train threshold in 742 episodes
Reached reward test threshold in 742 episodes
Reached reward train threshold in 743 episodes
Reached reward test threshold in 744 episodes
Reached reward train threshold in 745 episodes
Reached reward test threshold in 745 episodes
Reached reward train threshold in 746 episodes
Reached reward test threshold in 746 episodes
```

```
Reached reward train threshold in 747 episodes
Reached reward test threshold in 747 episodes
Reached reward train threshold in 748 episodes
Reached reward train threshold in 749 episodes
Reached reward test threshold in 749 episodes
| Episode: 750 | Mean Train Rewards:   200.1 | Mean Test Rewards:   211.5 |
Reached reward train threshold in 750 episodes
Reached reward test threshold in 751 episodes
Reached reward train threshold in 752 episodes
Reached reward test threshold in 752 episodes
Reached reward train threshold in 753 episodes
Reached reward test threshold in 753 episodes
Reached reward train threshold in 754 episodes
Reached reward train threshold in 755 episodes
Reached reward test threshold in 755 episodes
Reached reward train threshold in 756 episodes
Reached reward train threshold in 757 episodes
Reached reward test threshold in 757 episodes
Reached reward train threshold in 758 episodes
Reached reward test threshold in 758 episodes
Reached reward test threshold in 759 episodes
| Episode: 760 | Mean Train Rewards:   214.6 | Mean Test Rewards:   209.0 |
Reached reward train threshold in 760 episodes
Reached reward test threshold in 760 episodes
Reached reward train threshold in 761 episodes
Reached reward test threshold in 761 episodes
Reached reward train threshold in 762 episodes
Reached reward test threshold in 762 episodes
Reached reward train threshold in 763 episodes
Reached reward test threshold in 763 episodes
Reached reward train threshold in 764 episodes
Reached reward test threshold in 765 episodes
Reached reward train threshold in 766 episodes
Reached reward train threshold in 767 episodes
Reached reward test threshold in 767 episodes
Reached reward train threshold in 768 episodes
Reached reward train threshold in 769 episodes
Reached reward test threshold in 769 episodes
| Episode: 770 | Mean Train Rewards:   220.1 | Mean Test Rewards:   198.1 |
Reached reward train threshold in 770 episodes
Reached reward test threshold in 770 episodes
Reached reward train threshold in 771 episodes
Reached reward train threshold in 772 episodes
Reached reward test threshold in 772 episodes
Reached reward train threshold in 773 episodes
Reached reward test threshold in 773 episodes
Reached reward train threshold in 774 episodes
Reached reward test threshold in 774 episodes
Reached reward train threshold in 775 episodes
Reached reward test threshold in 775 episodes
Reached reward train threshold in 777 episodes
Reached reward test threshold in 777 episodes
Reached reward train threshold in 778 episodes
Reached reward test threshold in 778 episodes
Reached reward test threshold in 779 episodes
| Episode: 780 | Mean Train Rewards:   213.0 | Mean Test Rewards:   198.7 |
Reached reward test threshold in 780 episodes
Reached reward train threshold in 781 episodes
Reached reward test threshold in 781 episodes
Reached reward train threshold in 782 episodes
Reached reward train threshold in 783 episodes
Reached reward test threshold in 783 episodes
Reached reward train threshold in 784 episodes
Reached reward train threshold in 785 episodes
Reached reward test threshold in 785 episodes
Reached reward train threshold in 786 episodes
Reached reward test threshold in 786 episodes
Reached reward train threshold in 787 episodes
Reached reward test threshold in 787 episodes
```

```
Reached reward train threshold in 788 episodes
Reached reward test threshold in 788 episodes
Reached reward train threshold in 789 episodes
Reached reward test threshold in 789 episodes
| Episode: 790 | Mean Train Rewards:   221.0 | Mean Test Rewards:   203.6 |
Reached reward train threshold in 790 episodes
Reached reward test threshold in 790 episodes
Reached reward train threshold in 791 episodes
Reached reward test threshold in 791 episodes
Reached reward train threshold in 792 episodes
Reached reward test threshold in 792 episodes
Reached reward train threshold in 793 episodes
Reached reward test threshold in 793 episodes
Reached reward train threshold in 794 episodes
Reached reward train threshold in 795 episodes
Reached reward test threshold in 795 episodes
Reached reward train threshold in 796 episodes
Reached reward test threshold in 796 episodes
Reached reward train threshold in 797 episodes
Reached reward test threshold in 797 episodes
Reached reward train threshold in 798 episodes
Reached reward train threshold in 799 episodes
Reached reward test threshold in 799 episodes
| Episode: 800 | Mean Train Rewards:   224.7 | Mean Test Rewards:   220.1 |
Reached reward train threshold in 800 episodes
Reached reward test threshold in 800 episodes
Reached reward train threshold in 801 episodes
Reached reward test threshold in 801 episodes
Reached reward train threshold in 802 episodes
Reached reward test threshold in 802 episodes
Reached reward train threshold in 803 episodes
Reached reward test threshold in 803 episodes
Reached reward train threshold in 805 episodes
Reached reward test threshold in 806 episodes
Reached reward train threshold in 807 episodes
Reached reward test threshold in 807 episodes
Reached reward train threshold in 808 episodes
Reached reward train threshold in 809 episodes
| Episode: 810 | Mean Train Rewards:   213.5 | Mean Test Rewards:   208.9 |
Reached reward train threshold in 813 episodes
Reached reward test threshold in 813 episodes
Reached reward test threshold in 814 episodes
Reached reward train threshold in 815 episodes
Reached reward test threshold in 817 episodes
| Episode: 820 | Mean Train Rewards:   169.0 | Mean Test Rewards:   170.3 |
Reached reward train threshold in 820 episodes
Reached reward train threshold in 822 episodes
Reached reward train threshold in 823 episodes
Reached reward train threshold in 825 episodes
Reached reward test threshold in 828 episodes
| Episode: 830 | Mean Train Rewards:   144.0 | Mean Test Rewards:   164.4 |
Reached reward train threshold in 830 episodes
Reached reward train threshold in 831 episodes
Reached reward train threshold in 832 episodes
Reached reward train threshold in 833 episodes
Reached reward test threshold in 833 episodes
Reached reward train threshold in 834 episodes
Reached reward test threshold in 834 episodes
Reached reward train threshold in 835 episodes
Reached reward train threshold in 836 episodes
Reached reward test threshold in 836 episodes
Reached reward train threshold in 837 episodes
Reached reward train threshold in 838 episodes
Reached reward test threshold in 838 episodes
Reached reward train threshold in 839 episodes
Reached reward test threshold in 839 episodes
| Episode: 840 | Mean Train Rewards:   182.2 | Mean Test Rewards:   189.1 |
Reached reward train threshold in 840 episodes
Reached reward test threshold in 841 episodes
```

```
Reached reward test threshold in 842 episodes
Reached reward train threshold in 843 episodes
Reached reward test threshold in 843 episodes
Reached reward train threshold in 844 episodes
Reached reward train threshold in 846 episodes
Reached reward test threshold in 846 episodes
Reached reward train threshold in 847 episodes
Reached reward test threshold in 847 episodes
Reached reward test threshold in 848 episodes
Reached reward test threshold in 849 episodes
| Episode: 850 | Mean Train Rewards:   195.5 | Mean Test Rewards:   204.4 |
Reached reward train threshold in 850 episodes
Reached reward test threshold in 850 episodes
Reached reward train threshold in 851 episodes
Reached reward test threshold in 851 episodes
Reached reward test threshold in 852 episodes
Reached reward train threshold in 853 episodes
Reached reward test threshold in 853 episodes
Reached reward test threshold in 854 episodes
Reached reward train threshold in 855 episodes
Reached reward train threshold in 856 episodes
Reached reward test threshold in 856 episodes
Reached reward train threshold in 858 episodes
Reached reward test threshold in 858 episodes
Reached reward train threshold in 859 episodes
Reached reward test threshold in 859 episodes
| Episode: 860 | Mean Train Rewards:   186.6 | Mean Test Rewards:   212.0 |
Reached reward train threshold in 860 episodes
Reached reward test threshold in 860 episodes
Reached reward train threshold in 861 episodes
Reached reward train threshold in 862 episodes
Reached reward test threshold in 862 episodes
Reached reward train threshold in 863 episodes
Reached reward test threshold in 863 episodes
Reached reward train threshold in 864 episodes
Reached reward test threshold in 864 episodes
Reached reward train threshold in 865 episodes
Reached reward test threshold in 865 episodes
Reached reward train threshold in 866 episodes
Reached reward test threshold in 866 episodes
Reached reward train threshold in 867 episodes
Reached reward test threshold in 867 episodes
Reached reward train threshold in 868 episodes
Reached reward test threshold in 868 episodes
Reached reward train threshold in 869 episodes
| Episode: 870 | Mean Train Rewards:   208.5 | Mean Test Rewards:   214.3 |
Reached reward train threshold in 870 episodes
Reached reward test threshold in 870 episodes
Reached reward train threshold in 871 episodes
Reached reward test threshold in 871 episodes
Reached reward train threshold in 872 episodes
Reached reward test threshold in 872 episodes
Reached reward train threshold in 873 episodes
Reached reward test threshold in 873 episodes
Reached reward train threshold in 874 episodes
Reached reward test threshold in 874 episodes
Reached reward train threshold in 875 episodes
Reached reward test threshold in 875 episodes
Reached reward train threshold in 876 episodes
Reached reward test threshold in 876 episodes
Reached reward train threshold in 877 episodes
Reached reward test threshold in 877 episodes
Reached reward train threshold in 878 episodes
Reached reward test threshold in 878 episodes
Reached reward train threshold in 879 episodes
Reached reward test threshold in 879 episodes
| Episode: 880 | Mean Train Rewards:   241.0 | Mean Test Rewards:   217.5 |
Reached reward train threshold in 880 episodes
Reached reward test threshold in 880 episodes
```

```
Reached reward train threshold in 881 episodes
Reached reward test threshold in 881 episodes
Reached reward train threshold in 882 episodes
Reached reward train threshold in 883 episodes
Reached reward test threshold in 883 episodes
Reached reward train threshold in 884 episodes
Reached reward test threshold in 884 episodes
Reached reward train threshold in 885 episodes
Reached reward test threshold in 885 episodes
Reached reward train threshold in 886 episodes
Reached reward test threshold in 886 episodes
Reached reward train threshold in 887 episodes
Reached reward test threshold in 887 episodes
Reached reward test threshold in 888 episodes
Reached reward train threshold in 889 episodes
Reached reward test threshold in 889 episodes
| Episode: 890 | Mean Train Rewards:   249.7 | Mean Test Rewards:   226.1 |
Reached reward train threshold in 890 episodes
Reached reward test threshold in 890 episodes
Reached reward train threshold in 891 episodes
Reached reward train threshold in 892 episodes
Reached reward test threshold in 892 episodes
Reached reward train threshold in 893 episodes
Reached reward test threshold in 893 episodes
Reached reward train threshold in 894 episodes
Reached reward test threshold in 894 episodes
Reached reward train threshold in 895 episodes
Reached reward train threshold in 896 episodes
Reached reward test threshold in 896 episodes
Reached reward test threshold in 897 episodes
Reached reward train threshold in 898 episodes
Reached reward train threshold in 899 episodes
Reached reward test threshold in 899 episodes
| Episode: 900 | Mean Train Rewards:   235.0 | Mean Test Rewards:   224.3 |
Reached reward train threshold in 900 episodes
Reached reward test threshold in 900 episodes
Reached reward train threshold in 901 episodes
Reached reward test threshold in 901 episodes
Reached reward train threshold in 902 episodes
Reached reward test threshold in 902 episodes
Reached reward train threshold in 903 episodes
Reached reward test threshold in 903 episodes
Reached reward train threshold in 904 episodes
Reached reward test threshold in 904 episodes
Reached reward train threshold in 905 episodes
Reached reward test threshold in 905 episodes
Reached reward train threshold in 906 episodes
Reached reward train threshold in 907 episodes
Reached reward test threshold in 907 episodes
Reached reward train threshold in 908 episodes
Reached reward train threshold in 909 episodes
Reached reward test threshold in 909 episodes
| Episode: 910 | Mean Train Rewards:   231.2 | Mean Test Rewards:   215.4 |
Reached reward train threshold in 910 episodes
Reached reward test threshold in 910 episodes
Reached reward train threshold in 911 episodes
Reached reward train threshold in 912 episodes
Reached reward test threshold in 912 episodes
Reached reward train threshold in 913 episodes
Reached reward test threshold in 914 episodes
Reached reward train threshold in 915 episodes
Reached reward train threshold in 916 episodes
Reached reward test threshold in 916 episodes
Reached reward train threshold in 917 episodes
Reached reward test threshold in 917 episodes
Reached reward train threshold in 918 episodes
Reached reward train threshold in 919 episodes
Reached reward test threshold in 919 episodes
| Episode: 920 | Mean Train Rewards:   232.7 | Mean Test Rewards:   205.1 |
```

```
Reached reward train threshold in 920 episodes
Reached reward test threshold in 920 episodes
Reached reward train threshold in 921 episodes
Reached reward train threshold in 922 episodes
Reached reward test threshold in 922 episodes
Reached reward train threshold in 923 episodes
Reached reward test threshold in 923 episodes
Reached reward train threshold in 924 episodes
Reached reward train threshold in 925 episodes
Reached reward test threshold in 925 episodes
Reached reward train threshold in 926 episodes
Reached reward test threshold in 926 episodes
Reached reward train threshold in 927 episodes
Reached reward test threshold in 927 episodes
Reached reward train threshold in 928 episodes
Reached reward train threshold in 929 episodes
Reached reward test threshold in 929 episodes
| Episode: 930 | Mean Train Rewards:   242.0 | Mean Test Rewards:   201.1 |
Reached reward train threshold in 930 episodes
Reached reward test threshold in 930 episodes
Reached reward train threshold in 931 episodes
Reached reward train threshold in 932 episodes
Reached reward test threshold in 932 episodes
Reached reward train threshold in 933 episodes
Reached reward test threshold in 933 episodes
Reached reward train threshold in 934 episodes
Reached reward test threshold in 934 episodes
Reached reward train threshold in 935 episodes
Reached reward train threshold in 936 episodes
Reached reward test threshold in 936 episodes
Reached reward train threshold in 937 episodes
Reached reward test threshold in 937 episodes
Reached reward train threshold in 938 episodes
Reached reward test threshold in 938 episodes
Reached reward train threshold in 939 episodes
Reached reward test threshold in 939 episodes
| Episode: 940 | Mean Train Rewards:   242.1 | Mean Test Rewards:   216.6 |
Reached reward test threshold in 940 episodes
Reached reward train threshold in 941 episodes
Reached reward test threshold in 941 episodes
Reached reward train threshold in 942 episodes
Reached reward test threshold in 942 episodes
Reached reward train threshold in 943 episodes
Reached reward test threshold in 943 episodes
Reached reward train threshold in 944 episodes
Reached reward test threshold in 944 episodes
Reached reward train threshold in 945 episodes
Reached reward train threshold in 946 episodes
Reached reward test threshold in 946 episodes
Reached reward train threshold in 947 episodes
Reached reward test threshold in 947 episodes
Reached reward train threshold in 948 episodes
Reached reward test threshold in 948 episodes
Reached reward train threshold in 949 episodes
| Episode: 950 | Mean Train Rewards:   238.8 | Mean Test Rewards:   216.6 |
Reached reward train threshold in 951 episodes
Reached reward test threshold in 951 episodes
Reached reward train threshold in 952 episodes
Reached reward test threshold in 952 episodes
Reached reward train threshold in 953 episodes
Reached reward test threshold in 953 episodes
Reached reward train threshold in 955 episodes
Reached reward test threshold in 955 episodes
Reached reward train threshold in 956 episodes
Reached reward train threshold in 958 episodes
Reached reward test threshold in 958 episodes
Reached reward train threshold in 959 episodes
Reached reward test threshold in 959 episodes
| Episode: 960 | Mean Train Rewards:   205.2 | Mean Test Rewards:   213.0 |
```

```
Reached reward test threshold in 960 episodes
Reached reward train threshold in 961 episodes
Reached reward test threshold in 961 episodes
Reached reward train threshold in 962 episodes
Reached reward test threshold in 962 episodes
Reached reward train threshold in 963 episodes
Reached reward test threshold in 963 episodes
Reached reward train threshold in 964 episodes
Reached reward test threshold in 964 episodes
Reached reward train threshold in 965 episodes
Reached reward test threshold in 965 episodes
Reached reward train threshold in 966 episodes
Reached reward test threshold in 966 episodes
Reached reward train threshold in 967 episodes
Reached reward test threshold in 967 episodes
Reached reward test threshold in 968 episodes
Reached reward train threshold in 969 episodes
Reached reward test threshold in 969 episodes
| Episode: 970 | Mean Train Rewards:   187.8 | Mean Test Rewards:   211.7 |
Reached reward test threshold in 970 episodes
Reached reward train threshold in 971 episodes
Reached reward test threshold in 971 episodes
Reached reward train threshold in 972 episodes
Reached reward train threshold in 973 episodes
Reached reward test threshold in 973 episodes
Reached reward train threshold in 974 episodes
Reached reward train threshold in 975 episodes
Reached reward test threshold in 975 episodes
Reached reward train threshold in 976 episodes
Reached reward test threshold in 976 episodes
Reached reward train threshold in 977 episodes
Reached reward test threshold in 977 episodes
Reached reward train threshold in 978 episodes
Reached reward test threshold in 978 episodes
Reached reward train threshold in 979 episodes
Reached reward test threshold in 979 episodes
| Episode: 980 | Mean Train Rewards:   213.6 | Mean Test Rewards:   215.2 |
Reached reward train threshold in 980 episodes
Reached reward test threshold in 980 episodes
Reached reward train threshold in 981 episodes
Reached reward test threshold in 981 episodes
Reached reward train threshold in 982 episodes
Reached reward test threshold in 982 episodes
Reached reward train threshold in 983 episodes
Reached reward test threshold in 983 episodes
Reached reward train threshold in 984 episodes
Reached reward test threshold in 984 episodes
Reached reward train threshold in 985 episodes
Reached reward test threshold in 985 episodes
Reached reward train threshold in 986 episodes
Reached reward test threshold in 986 episodes
Reached reward train threshold in 987 episodes
Reached reward test threshold in 987 episodes
Reached reward train threshold in 988 episodes
Reached reward test threshold in 988 episodes
Reached reward train threshold in 989 episodes
Reached reward test threshold in 989 episodes
| Episode: 990 | Mean Train Rewards:   234.3 | Mean Test Rewards:   229.2 |
Reached reward train threshold in 990 episodes
Reached reward test threshold in 990 episodes
Reached reward train threshold in 991 episodes
Reached reward test threshold in 991 episodes
Reached reward train threshold in 992 episodes
Reached reward test threshold in 992 episodes
Reached reward train threshold in 993 episodes
Reached reward test threshold in 993 episodes
Reached reward train threshold in 994 episodes
Reached reward test threshold in 994 episodes
Reached reward train threshold in 995 episodes
```

```
Reached reward test threshold in 995 episodes
Reached reward test threshold in 996 episodes
Reached reward train threshold in 997 episodes
Reached reward train threshold in 998 episodes
Reached reward test threshold in 998 episodes
Reached reward train threshold in 999 episodes
Reached reward test threshold in 999 episodes
| Episode: 1000 | Mean Train Rewards:   250.3 | Mean Test Rewards:   237.4 |
Reached reward train threshold in 1000 episodes
Reached reward test threshold in 1000 episodes
Reached reward train threshold in 1001 episodes
Reached reward test threshold in 1001 episodes
Reached reward train threshold in 1002 episodes
Reached reward test threshold in 1002 episodes
Reached reward train threshold in 1003 episodes
Reached reward test threshold in 1003 episodes
Reached reward train threshold in 1004 episodes
Reached reward test threshold in 1004 episodes
Reached reward train threshold in 1005 episodes
Reached reward train threshold in 1006 episodes
Reached reward test threshold in 1006 episodes
Reached reward train threshold in 1007 episodes
Reached reward test threshold in 1007 episodes
Reached reward train threshold in 1008 episodes
Reached reward test threshold in 1008 episodes
Reached reward train threshold in 1009 episodes
Reached reward test threshold in 1009 episodes
| Episode: 1010 | Mean Train Rewards:   249.0 | Mean Test Rewards:   237.8 |
Reached reward train threshold in 1010 episodes
Reached reward test threshold in 1010 episodes
Reached reward train threshold in 1011 episodes
Reached reward test threshold in 1011 episodes
Reached reward train threshold in 1012 episodes
Reached reward train threshold in 1013 episodes
Reached reward test threshold in 1013 episodes
Reached reward test threshold in 1014 episodes
Reached reward train threshold in 1015 episodes
Reached reward test threshold in 1015 episodes
Reached reward train threshold in 1016 episodes
Reached reward test threshold in 1016 episodes
Reached reward train threshold in 1017 episodes
Reached reward test threshold in 1017 episodes
Reached reward train threshold in 1018 episodes
Reached reward test threshold in 1018 episodes
Reached reward train threshold in 1019 episodes
Reached reward test threshold in 1019 episodes
| Episode: 1020 | Mean Train Rewards:   252.6 | Mean Test Rewards:   224.0 |
Reached reward train threshold in 1020 episodes
Reached reward train threshold in 1021 episodes
Reached reward test threshold in 1021 episodes
Reached reward train threshold in 1022 episodes
Reached reward test threshold in 1022 episodes
Reached reward train threshold in 1023 episodes
Reached reward test threshold in 1023 episodes
Reached reward test threshold in 1024 episodes
Reached reward train threshold in 1025 episodes
Reached reward test threshold in 1025 episodes
Reached reward train threshold in 1026 episodes
Reached reward test threshold in 1026 episodes
Reached reward test threshold in 1027 episodes
Reached reward train threshold in 1028 episodes
Reached reward test threshold in 1028 episodes
Reached reward train threshold in 1029 episodes
Reached reward test threshold in 1029 episodes
| Episode: 1030 | Mean Train Rewards:   242.3 | Mean Test Rewards:   220.8 |
Reached reward train threshold in 1030 episodes
Reached reward test threshold in 1030 episodes
Reached reward train threshold in 1031 episodes
Reached reward test threshold in 1031 episodes
```

```
Reached reward train threshold in 1032 episodes
Reached reward test threshold in 1032 episodes
Reached reward train threshold in 1033 episodes
Reached reward test threshold in 1033 episodes
Reached reward train threshold in 1034 episodes
Reached reward test threshold in 1034 episodes
Reached reward train threshold in 1035 episodes
Reached reward test threshold in 1035 episodes
Reached reward train threshold in 1036 episodes
Reached reward test threshold in 1036 episodes
Reached reward test threshold in 1037 episodes
Reached reward train threshold in 1038 episodes
Reached reward test threshold in 1038 episodes
Reached reward train threshold in 1039 episodes
Reached reward test threshold in 1039 episodes
| Episode: 1040 | Mean Train Rewards:   227.6 | Mean Test Rewards:   228.6 |
Reached reward train threshold in 1040 episodes
Reached reward test threshold in 1040 episodes
Reached reward train threshold in 1041 episodes
Reached reward test threshold in 1041 episodes
Reached reward train threshold in 1042 episodes
Reached reward test threshold in 1042 episodes
Reached reward train threshold in 1043 episodes
Reached reward test threshold in 1043 episodes
Reached reward test threshold in 1044 episodes
Reached reward train threshold in 1045 episodes
Reached reward test threshold in 1045 episodes
Reached reward train threshold in 1046 episodes
Reached reward test threshold in 1046 episodes
Reached reward train threshold in 1047 episodes
Reached reward test threshold in 1047 episodes
Reached reward test threshold in 1048 episodes
Reached reward train threshold in 1049 episodes
Reached reward test threshold in 1049 episodes
| Episode: 1050 | Mean Train Rewards:   205.3 | Mean Test Rewards:   229.0 |
Reached reward train threshold in 1050 episodes
Reached reward test threshold in 1050 episodes
Reached reward train threshold in 1051 episodes
Reached reward test threshold in 1051 episodes
Reached reward train threshold in 1052 episodes
Reached reward test threshold in 1052 episodes
Reached reward train threshold in 1053 episodes
Reached reward test threshold in 1053 episodes
Reached reward train threshold in 1054 episodes
Reached reward test threshold in 1054 episodes
Reached reward train threshold in 1055 episodes
Reached reward test threshold in 1055 episodes
Reached reward train threshold in 1056 episodes
Reached reward test threshold in 1056 episodes
Reached reward train threshold in 1057 episodes
Reached reward train threshold in 1058 episodes
Reached reward test threshold in 1058 episodes
Reached reward train threshold in 1059 episodes
Reached reward test threshold in 1059 episodes
| Episode: 1060 | Mean Train Rewards:   214.5 | Mean Test Rewards:   226.1 |
Reached reward train threshold in 1060 episodes
Reached reward test threshold in 1060 episodes
Reached reward train threshold in 1061 episodes
Reached reward test threshold in 1061 episodes
Reached reward train threshold in 1062 episodes
Reached reward test threshold in 1062 episodes
Reached reward train threshold in 1063 episodes
Reached reward test threshold in 1063 episodes
Reached reward test threshold in 1064 episodes
Reached reward test threshold in 1065 episodes
Reached reward train threshold in 1066 episodes
Reached reward test threshold in 1066 episodes
Reached reward train threshold in 1067 episodes
Reached reward train threshold in 1068 episodes
```

```
Reached reward test threshold in 1068 episodes
Reached reward train threshold in 1069 episodes
| Episode: 1070 | Mean Train Rewards:   219.8 | Mean Test Rewards:   213.5 |
Reached reward train threshold in 1070 episodes
Reached reward test threshold in 1070 episodes
Reached reward train threshold in 1071 episodes
Reached reward test threshold in 1071 episodes
Reached reward train threshold in 1072 episodes
Reached reward test threshold in 1072 episodes
Reached reward train threshold in 1073 episodes
Reached reward test threshold in 1073 episodes
Reached reward train threshold in 1074 episodes
Reached reward test threshold in 1074 episodes
Reached reward train threshold in 1075 episodes
Reached reward train threshold in 1076 episodes
Reached reward test threshold in 1076 episodes
Reached reward train threshold in 1077 episodes
Reached reward test threshold in 1077 episodes
Reached reward train threshold in 1078 episodes
Reached reward test threshold in 1078 episodes
Reached reward train threshold in 1079 episodes
Reached reward test threshold in 1079 episodes
| Episode: 1080 | Mean Train Rewards:   228.7 | Mean Test Rewards:   214.5 |
Reached reward train threshold in 1080 episodes
Reached reward train threshold in 1081 episodes
Reached reward test threshold in 1081 episodes
Reached reward train threshold in 1082 episodes
Reached reward test threshold in 1083 episodes
Reached reward train threshold in 1084 episodes
Reached reward test threshold in 1084 episodes
Reached reward train threshold in 1085 episodes
Reached reward train threshold in 1086 episodes
Reached reward test threshold in 1086 episodes
Reached reward train threshold in 1087 episodes
Reached reward test threshold in 1087 episodes
Reached reward train threshold in 1088 episodes
Reached reward test threshold in 1088 episodes
Reached reward train threshold in 1089 episodes
Reached reward test threshold in 1089 episodes
| Episode: 1090 | Mean Train Rewards:   234.1 | Mean Test Rewards:   216.3 |
Reached reward train threshold in 1090 episodes
Reached reward test threshold in 1090 episodes
Reached reward train threshold in 1091 episodes
Reached reward test threshold in 1091 episodes
Reached reward train threshold in 1092 episodes
Reached reward test threshold in 1092 episodes
Reached reward train threshold in 1093 episodes
Reached reward test threshold in 1093 episodes
Reached reward train threshold in 1094 episodes
Reached reward test threshold in 1094 episodes
Reached reward train threshold in 1095 episodes
Reached reward test threshold in 1095 episodes
Reached reward train threshold in 1096 episodes
Reached reward test threshold in 1096 episodes
Reached reward train threshold in 1097 episodes
Reached reward test threshold in 1097 episodes
Reached reward train threshold in 1098 episodes
Reached reward test threshold in 1098 episodes
Reached reward train threshold in 1099 episodes
Reached reward test threshold in 1099 episodes
| Episode: 1100 | Mean Train Rewards:   238.0 | Mean Test Rewards:   235.9 |
Reached reward train threshold in 1100 episodes
Reached reward test threshold in 1100 episodes
Reached reward train threshold in 1101 episodes
Reached reward test threshold in 1101 episodes
Reached reward test threshold in 1102 episodes
Reached reward train threshold in 1103 episodes
Reached reward test threshold in 1103 episodes
Reached reward train threshold in 1104 episodes
```

```
Reached reward test threshold in 1104 episodes
Reached reward train threshold in 1105 episodes
Reached reward train threshold in 1106 episodes
Reached reward test threshold in 1106 episodes
Reached reward train threshold in 1107 episodes
Reached reward test threshold in 1107 episodes
Reached reward train threshold in 1108 episodes
Reached reward test threshold in 1108 episodes
Reached reward train threshold in 1109 episodes
| Episode: 1110 | Mean Train Rewards:   237.0 | Mean Test Rewards:   230.7 |
Reached reward train threshold in 1110 episodes
Reached reward test threshold in 1110 episodes
Reached reward train threshold in 1111 episodes
Reached reward train threshold in 1112 episodes
Reached reward test threshold in 1112 episodes
Reached reward train threshold in 1113 episodes
Reached reward test threshold in 1113 episodes
Reached reward train threshold in 1114 episodes
Reached reward train threshold in 1115 episodes
Reached reward test threshold in 1115 episodes
Reached reward train threshold in 1116 episodes
Reached reward test threshold in 1116 episodes
Reached reward train threshold in 1117 episodes
Reached reward test threshold in 1117 episodes
Reached reward train threshold in 1118 episodes
Reached reward train threshold in 1119 episodes
Reached reward test threshold in 1119 episodes
| Episode: 1120 | Mean Train Rewards:   235.7 | Mean Test Rewards:   210.5 |
Reached reward train threshold in 1120 episodes
Reached reward test threshold in 1120 episodes
Reached reward train threshold in 1121 episodes
Reached reward train threshold in 1122 episodes
Reached reward test threshold in 1122 episodes
Reached reward train threshold in 1123 episodes
Reached reward test threshold in 1123 episodes
Reached reward test threshold in 1124 episodes
Reached reward train threshold in 1125 episodes
Reached reward test threshold in 1125 episodes
Reached reward train threshold in 1126 episodes
Reached reward test threshold in 1126 episodes
Reached reward train threshold in 1127 episodes
Reached reward test threshold in 1127 episodes
Reached reward train threshold in 1128 episodes
Reached reward test threshold in 1128 episodes
Reached reward train threshold in 1129 episodes
| Episode: 1130 | Mean Train Rewards:   237.8 | Mean Test Rewards:   210.5 |
Reached reward train threshold in 1130 episodes
Reached reward test threshold in 1130 episodes
Reached reward train threshold in 1131 episodes
Reached reward test threshold in 1131 episodes
Reached reward train threshold in 1132 episodes
Reached reward train threshold in 1133 episodes
Reached reward test threshold in 1134 episodes
Reached reward train threshold in 1135 episodes
Reached reward test threshold in 1135 episodes
Reached reward train threshold in 1136 episodes
Reached reward test threshold in 1136 episodes
Reached reward train threshold in 1137 episodes
Reached reward test threshold in 1137 episodes
Reached reward train threshold in 1138 episodes
Reached reward test threshold in 1138 episodes
Reached reward train threshold in 1139 episodes
| Episode: 1140 | Mean Train Rewards:   232.8 | Mean Test Rewards:   210.3 |
Reached reward train threshold in 1140 episodes
Reached reward test threshold in 1140 episodes
Reached reward train threshold in 1141 episodes
Reached reward test threshold in 1141 episodes
Reached reward train threshold in 1142 episodes
Reached reward test threshold in 1142 episodes
```

```
Reached reward train threshold in 1143 episodes
Reached reward test threshold in 1143 episodes
Reached reward train threshold in 1144 episodes
Reached reward train threshold in 1146 episodes
Reached reward test threshold in 1146 episodes
Reached reward train threshold in 1147 episodes
Reached reward test threshold in 1147 episodes
Reached reward train threshold in 1148 episodes
Reached reward train threshold in 1149 episodes
Reached reward test threshold in 1149 episodes
| Episode: 1150 | Mean Train Rewards:   232.7 | Mean Test Rewards:   201.4 |
Reached reward train threshold in 1150 episodes
Reached reward test threshold in 1150 episodes
Reached reward test threshold in 1154 episodes
Reached reward train threshold in 1155 episodes
Reached reward test threshold in 1155 episodes
Reached reward train threshold in 1156 episodes
Reached reward test threshold in 1157 episodes
Reached reward train threshold in 1158 episodes
Reached reward test threshold in 1158 episodes
Reached reward train threshold in 1159 episodes
Reached reward test threshold in 1159 episodes
| Episode: 1160 | Mean Train Rewards:   191.0 | Mean Test Rewards:   162.7 |
Reached reward train threshold in 1161 episodes
Reached reward test threshold in 1161 episodes
Reached reward train threshold in 1162 episodes
Reached reward train threshold in 1163 episodes
Reached reward train threshold in 1164 episodes
Reached reward test threshold in 1164 episodes
Reached reward train threshold in 1165 episodes
Reached reward test threshold in 1165 episodes
Reached reward train threshold in 1166 episodes
Reached reward train threshold in 1167 episodes
Reached reward test threshold in 1167 episodes
Reached reward train threshold in 1168 episodes
Reached reward test threshold in 1168 episodes
Reached reward train threshold in 1169 episodes
Reached reward test threshold in 1169 episodes
| Episode: 1170 | Mean Train Rewards:   199.7 | Mean Test Rewards:   158.5 |
Reached reward train threshold in 1170 episodes
Reached reward test threshold in 1170 episodes
Reached reward train threshold in 1171 episodes
Reached reward train threshold in 1172 episodes
Reached reward test threshold in 1172 episodes
Reached reward train threshold in 1173 episodes
Reached reward train threshold in 1174 episodes
Reached reward train threshold in 1175 episodes
Reached reward test threshold in 1175 episodes
Reached reward train threshold in 1176 episodes
Reached reward test threshold in 1176 episodes
Reached reward train threshold in 1177 episodes
Reached reward test threshold in 1177 episodes
Reached reward test threshold in 1178 episodes
Reached reward test threshold in 1179 episodes
| Episode: 1180 | Mean Train Rewards:   211.7 | Mean Test Rewards:   171.3 |
Reached reward train threshold in 1180 episodes
Reached reward test threshold in 1180 episodes
Reached reward train threshold in 1181 episodes
Reached reward test threshold in 1181 episodes
Reached reward test threshold in 1183 episodes
Reached reward train threshold in 1184 episodes
Reached reward test threshold in 1184 episodes
Reached reward train threshold in 1185 episodes
Reached reward test threshold in 1185 episodes
Reached reward train threshold in 1186 episodes
Reached reward test threshold in 1186 episodes
Reached reward train threshold in 1187 episodes
Reached reward test threshold in 1187 episodes
Reached reward train threshold in 1189 episodes
```

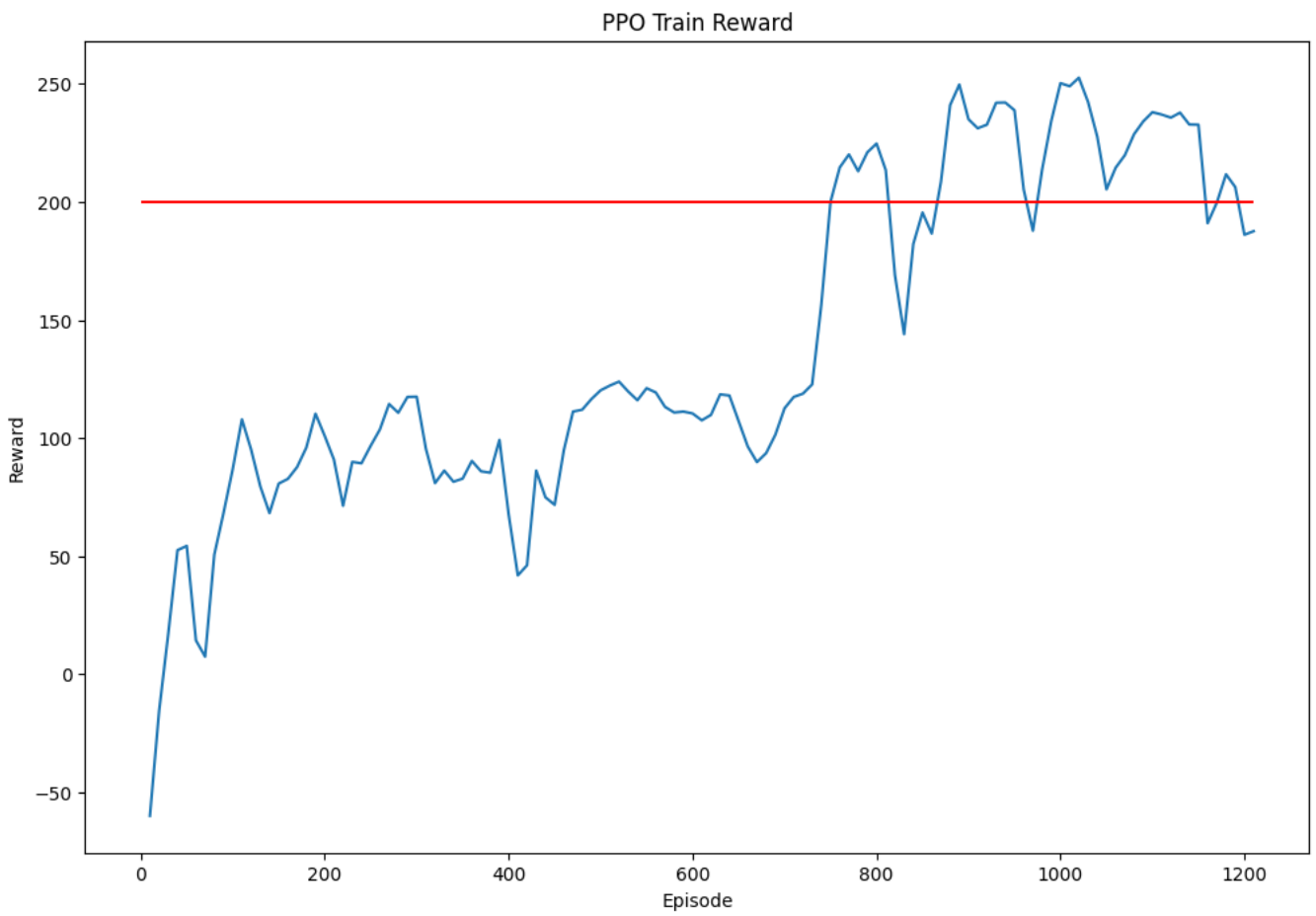
```
Reached reward test threshold in 1189 episodes
| Episode: 1190 | Mean Train Rewards:   206.3 | Mean Test Rewards:   183.2 |
Reached reward train threshold in 1190 episodes
Reached reward test threshold in 1190 episodes
Reached reward test threshold in 1191 episodes
Reached reward train threshold in 1192 episodes
Reached reward test threshold in 1192 episodes
Reached reward train threshold in 1193 episodes
Reached reward train threshold in 1195 episodes
Reached reward test threshold in 1196 episodes
Reached reward train threshold in 1197 episodes
Reached reward test threshold in 1197 episodes
Reached reward test threshold in 1199 episodes
| Episode: 1200 | Mean Train Rewards:   186.1 | Mean Test Rewards:   172.7 |
Reached reward train threshold in 1200 episodes
Reached reward train threshold in 1203 episodes
Reached reward train threshold in 1204 episodes
Reached reward train threshold in 1205 episodes
Reached reward train threshold in 1206 episodes
Reached reward test threshold in 1209 episodes
| Episode: 1210 | Mean Train Rewards:   187.6 | Mean Test Rewards:   130.2 |
Reached reward train threshold in 1210 episodes
Reached reward train threshold in 1211 episodes
Reached reward test threshold in 1211 episodes
Reached reward train threshold in 1212 episodes
Reached reward train threshold in 1213 episodes
```

```
---------------------------------------------------------------------------
RuntimeError                              Traceback (most recent call last)
File c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\torch\serialization.py:423,
in save(obj, f, pickle_module, pickle_protocol, _use_new_zipfile_serialization)
    422 with _open_zipfile_writer(f) as opened_zipfile:
--> 423     _save(obj, opened_zipfile, pickle_module, pickle_protocol)
    424     return

File c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\torch\serialization.py:650,
in _save(obj, zip_file, pickle_module, pickle_protocol)
    649 num_bytes = storage.nbytes()
--> 650 zip_file.write_record(name, storage.data_ptr(), num_bytes)

RuntimeError: [enforce fail at C:\actions-runner\_work\pytorch\pytorch\builder\windows\pytorc
h\caffe2\serialize\inline_container.cc:450] . PytorchStreamWriter failed writing file data/0:
file write failed

During handling of the above exception, another exception occurred:

RuntimeError                              Traceback (most recent call last)
Cell In[31], line 30
     28 if train_reward >= REWARD_THRESHOLD:
     29     print(f'Reached reward train threshold in {episode} episodes')
---> 30     torch.save(ppo_policy.state_dict(),
     31                f"./checkpoints/PPO/PPOTrain-{episode}.pth")
     33 if test_reward >= REWARD_THRESHOLD:
     34     print(f'Reached reward test threshold in {episode} episodes')

File c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\torch\serialization.py:424,
in save(obj, f, pickle_module, pickle_protocol, _use_new_zipfile_serialization)
    422     with _open_zipfile_writer(f) as opened_zipfile:
    423         _save(obj, opened_zipfile, pickle_module, pickle_protocol)
--> 424         return
    425 else:
    426     with _open_file_like(f, 'wb') as opened_file:

File c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\torch\serialization.py:290,
in _open_zipfile_writer_file.__exit__(self, *args)
    289 def __exit__(self, *args) -> None:
--> 290     self.file_like.write_end_of_file()

RuntimeError: [enforce fail at C:\actions-runner\_work\pytorch\pytorch\builder\windows\pytorc
h\caffe2\serialize\inline_container.cc:325] . unexpected pos 1984 vs 1921
```

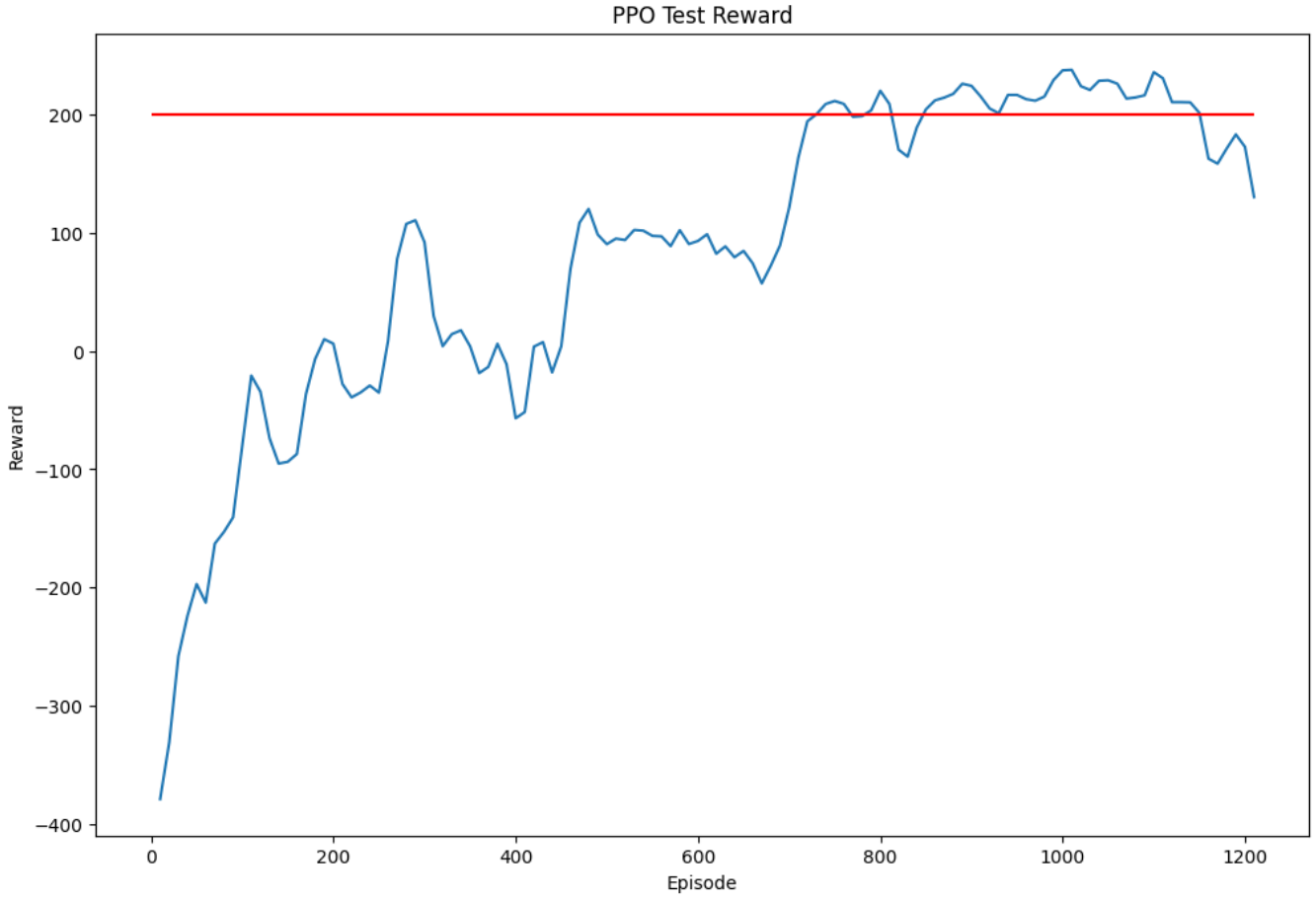
Evaluation of PPO Model

In [ ]:
```
ppo_train_rewards = np.array(ppo_train_rewards)
ppo_test_rewards = np.array(ppo_test_rewards)
```

In [ ]:
```
plt.figure(figsize=(12, 8))
plt.plot(x ,ppo_train_rewards, label='Train Reward')
plt.xlabel('Episode')
plt.ylabel('Reward')
plt.hlines(REWARD_THRESHOLD, 0, 1210, color='r')
plt.title("PPO Train Reward")
plt.show()
```



In [ ]:
```
plt.figure(figsize=(12,8))
plt.plot(x ,ppo_test_rewards, label='Test Reward')
plt.xlabel('Episode')
plt.ylabel('Reward')
plt.hlines(REWARD_THRESHOLD, 0, 1210, color='r')
plt.title("PPO Test Reward")
plt.show()
```



In [ ]:
```
plt.figure(figsize=(12,8))
plt.plot(x, ppo_train_rewards, label='Train Reward')
plt.plot(x, ppo_test_rewards, label='Test Reward')
plt.xlabel('Episode')
plt.ylabel('Reward')
plt.hlines(REWARD_THRESHOLD, 0, 1210, color='r')
plt.title("PPO Train Test Reward")
plt.legend(loc='lower right')
plt.show()
```

In [ ]:
```
print(
    f"Highest Train Episode: {np.argmax(ppo_train_rewards) + 1} Score: {np.max(ppo_train_rewa
print(
    f"Highest Test Episode: {np.argmax(ppo_test_rewards) + 1} Score: {np.max(ppo_test_rewards
```

```
Highest Train Episode: 1021 Score: 252.6
Highest Test Episode: 1011 Score: 237.8
```

Visualizing model

We will be choosing the values where both training and testing rewards are the higher than the reward threshold. Train >= 250 and Test >= 245

In [ ]:
```
best_episodes = np.intersect1d(np.where(
    np.array(ppo_train_rewards) >= 250), np.where(np.array(ppo_test_rewards) >= 235)) + 1
best_episodes
```

Out[ ]:
```
array([1011], dtype=int64)
```

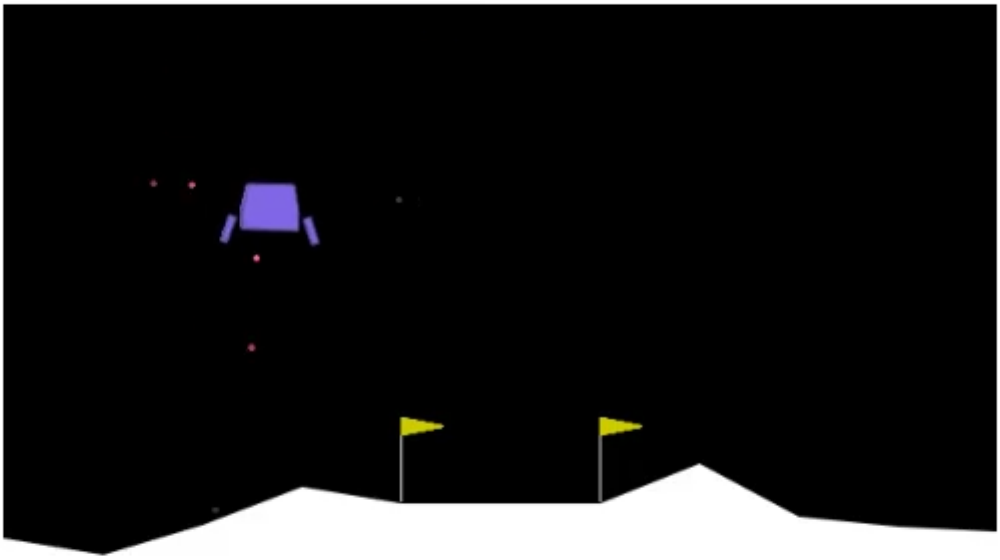
In [ ]:
```
best_episode = best_episodes[0]
ppo_policy.load_state_dict(torch.load(
    f"./checkpoints/PPO/PPOTrain-{best_episode}.pth"))
ppo_policy
```

Out[ ]:
```
ActorCritic(
  (actor): MLP(
    (net): Sequential(
      (0): Linear(in_features=8, out_features=128, bias=True)
      (1): Dropout(p=0.1, inplace=False)
      (2): PReLU(num_parameters=1)
      (3): Linear(in_features=128, out_features=128, bias=True)
      (4): Dropout(p=0.1, inplace=False)
      (5): PReLU(num_parameters=1)
      (6): Linear(in_features=128, out_features=4, bias=True)
    )
  )
  (critic): MLP(
    (net): Sequential(
      (0): Linear(in_features=8, out_features=128, bias=True)
      (1): Dropout(p=0.1, inplace=False)
      (2): PReLU(num_parameters=1)
      (3): Linear(in_features=128, out_features=128, bias=True)
      (4): Dropout(p=0.1, inplace=False)
      (5): PReLU(num_parameters=1)
      (6): Linear(in_features=128, out_features=1, bias=True)
    )
  )
)
```

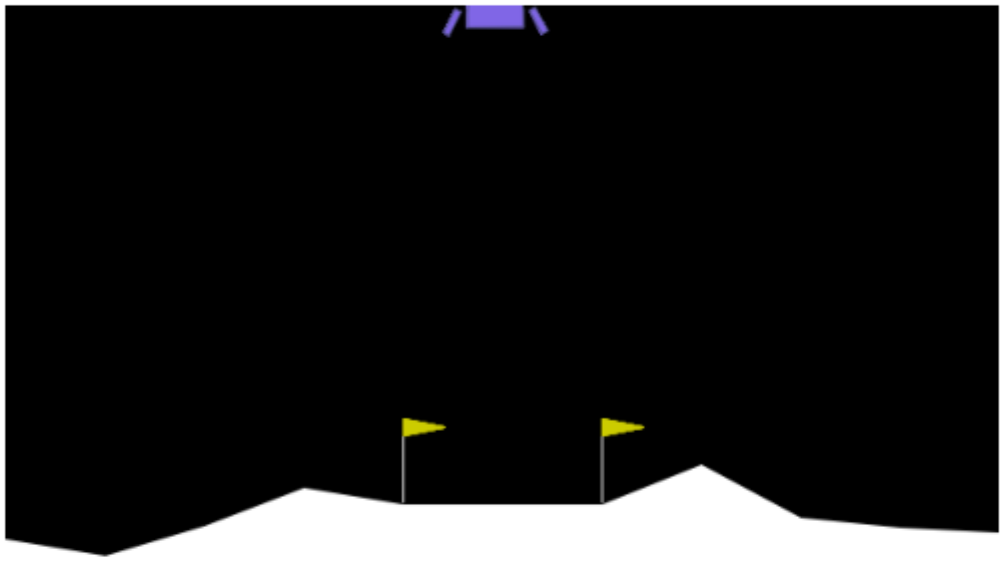
In [ ]:
```python
def show_ppo(env, policy, weights_path, iteration=1):
    policy.load_state_dict(torch.load(weights_path))
    policy.eval()

    rewards = []
    done = False
    episode_reward = 0
    frames = []
    state = env.reset()
    while not done:

        state = torch.FloatTensor(state).unsqueeze(0)

        with torch.no_grad():

            action_pred, _ = policy(state)

            action_prob = F.softmax(action_pred, dim=-1)

        action = torch.argmax(action_prob, dim=-1)

        state, reward, done, _ = env.step(action.item())
        screen = env.render(mode='rgb_array')
        episode_reward += reward
        frames.append(screen)
    return episode_reward, frames
```

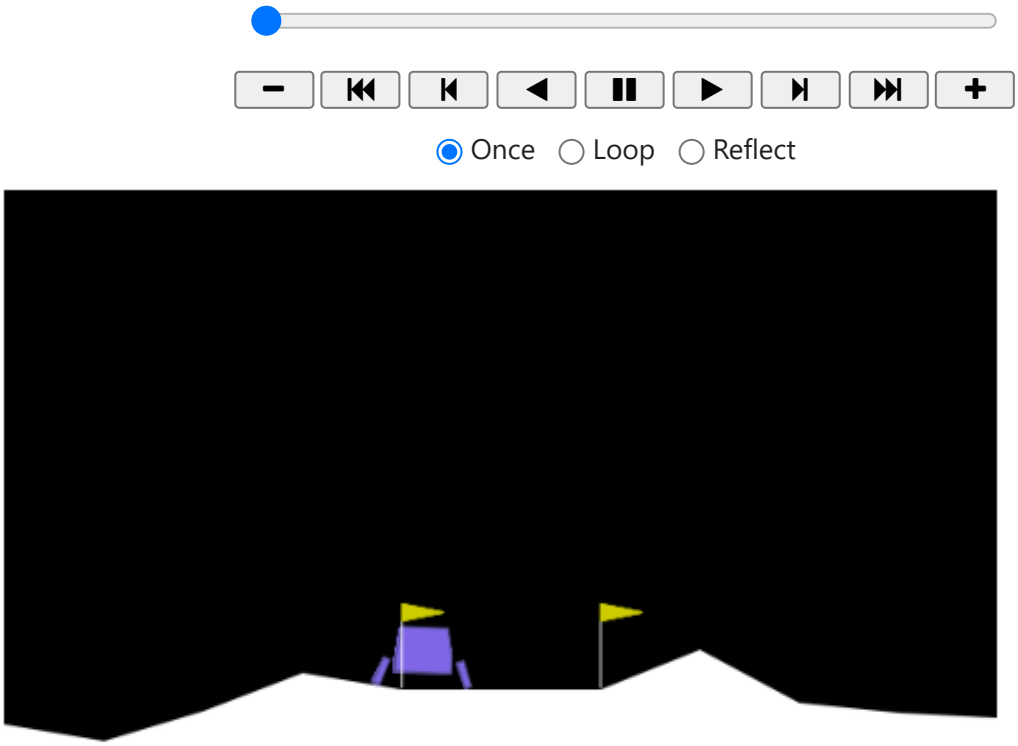
In [ ]:
```python
show_env = gym.make('LunarLander-v2',
                    continuous=False,
                    gravity=-10.0,
                    enable_wind=False,
                    wind_power=15.0,
                    turbulence_power=1.5
                    )
show_env.seed(75)
reward, frames = show_ppo(
    show_env, ppo_policy,  f"./checkpoints/PPO/PPOTrain-{best_episode}.pth")

print(f"Model Score: {reward}")

create_animation(frames, f"./videos/PPOTrain-{best_episode}-{reward}.mp4")
```

```
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:317: DeprecationWarning: WARN: Initializing wrapper in old step API which returns one bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\wrappers\step_api_compatibility.py:39: DeprecationWarning: WARN: Initializing environment in old step API which returns one bool instead of two. It is recommended to set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:256: DeprecationWarning: WARN: Function `env.seed(seed)` is marked as deprecated and will be removed in the future. Please use `env.reset(seed=seed)` instead.
  deprecation(
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\utils\passive_env_checker.py:241: DeprecationWarning: `np.bool8` is a deprecated alias for `np.bool_`.  (Deprecated NumPy 1.24)
  if not isinstance(terminated, (bool, np.bool8)):
c:\Users\Soh Hong Yu\anaconda3\envs\myenv\lib\site-packages\gym\core.py:43: DeprecationWarning: WARN: The argument mode in render method is deprecated; use render_mode during environment initialization instead.
See here for more information: https://www.gymlibrary.ml/content/api/
  deprecation(
Model Score: 258.4892892114275
```



Out[ ]:





Summary

A2C is a better model compared to DQN because it separates the learning of the policy and the value function into two separate neural networks. This allows the algorithm to directly optimize the policy by maximizing the expected cumulative reward, rather than estimating the Q-value function as in DQN. This can result in a more efficient learning process and a better exploration-exploitation trade-off. Another advantage of A2C is that it can handle continuous action spaces, whereas DQN is limited to discrete

action spaces. In addition, A2C can handle both on-policy and off-policy learning, while DQN is mainly used for off-policy learning. However, it should be noted that DQN can still perform well in some problems and is generally easier to implement and understand compared to A2C. Therefore our final model in this case will be A2C as it has the best performance