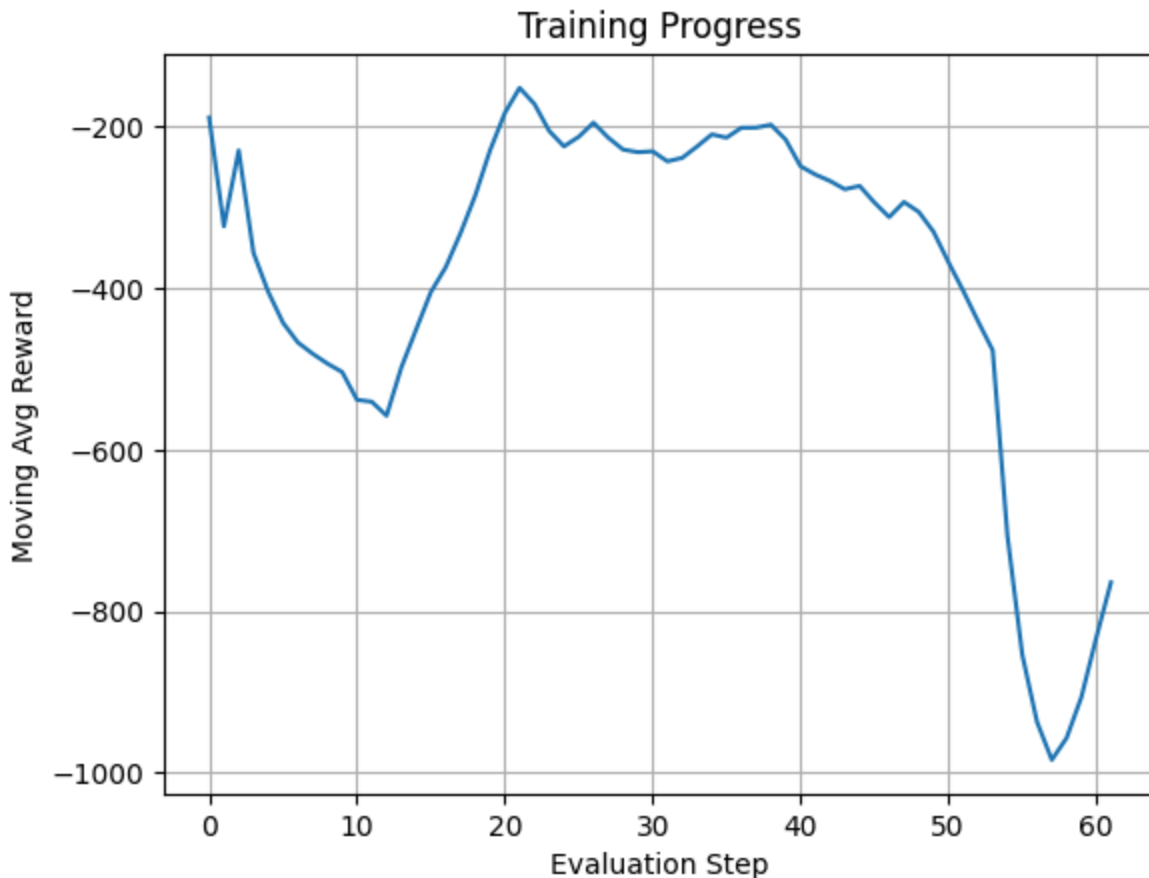


=== Running GRPO on LunarLander-v2 ===

Steps 16384: Avg Reward = -189.03, Moving Avg (10) = -189.03
Steps 32768: Avg Reward = -458.15, Moving Avg (10) = -323.59
Steps 49152: Avg Reward = -41.96, Moving Avg (10) = -229.71
Steps 65536: Avg Reward = -737.04, Moving Avg (10) = -356.55
Steps 81920: Avg Reward = -598.08, Moving Avg (10) = -404.85
Steps 98304: Avg Reward = -634.11, Moving Avg (10) = -443.06
Steps 114688: Avg Reward = -612.82, Moving Avg (10) = -467.31
Steps 131072: Avg Reward = -578.74, Moving Avg (10) = -481.24
Steps 147456: Avg Reward = -591.28, Moving Avg (10) = -493.47
Steps 163840: Avg Reward = -598.51, Moving Avg (10) = -503.97
Steps 180224: Avg Reward = -532.31, Moving Avg (10) = -538.30
Steps 196608: Avg Reward = -482.04, Moving Avg (10) = -540.69
Steps 212992: Avg Reward = -214.55, Moving Avg (10) = -557.95
Steps 229376: Avg Reward = -140.64, Moving Avg (10) = -498.31
Steps 245760: Avg Reward = -128.90, Moving Avg (10) = -451.39
Steps 262144: Avg Reward = -169.46, Moving Avg (10) = -404.93
Steps 278528: Avg Reward = -306.89, Moving Avg (10) = -374.33
Steps 294912: Avg Reward = -158.60, Moving Avg (10) = -332.32
Steps 311296: Avg Reward = -122.84, Moving Avg (10) = -285.47
Steps 327680: Avg Reward = -41.14, Moving Avg (10) = -229.74
Steps 344064: Avg Reward = -70.26, Moving Avg (10) = -183.53
Steps 360448: Avg Reward = -170.05, Moving Avg (10) = -152.33
Steps 376832: Avg Reward = -413.21, Moving Avg (10) = -172.20
Steps 393216: Avg Reward = -474.76, Moving Avg (10) = -205.61
Steps 409600: Avg Reward = -320.43, Moving Avg (10) = -224.77
Steps 425984: Avg Reward = -50.88, Moving Avg (10) = -212.91
Steps 442368: Avg Reward = -132.25, Moving Avg (10) = -195.44
Steps 458752: Avg Reward = -344.29, Moving Avg (10) = -214.01
Steps 475136: Avg Reward = -269.95, Moving Avg (10) = -228.72
Steps 491520: Avg Reward = -73.13, Moving Avg (10) = -231.92
Steps 507904: Avg Reward = -60.02, Moving Avg (10) = -230.90
Steps 524288: Avg Reward = -293.94, Moving Avg (10) = -243.29
Steps 540672: Avg Reward = -370.13, Moving Avg (10) = -238.98
Steps 557056: Avg Reward = -334.68, Moving Avg (10) = -224.97
Steps 573440: Avg Reward = -169.76, Moving Avg (10) = -209.90
Steps 589824: Avg Reward = -91.26, Moving Avg (10) = -213.94
Steps 606208: Avg Reward = -9.95, Moving Avg (10) = -201.71
Steps 622592: Avg Reward = -342.93, Moving Avg (10) = -201.57
Steps 638976: Avg Reward = -233.80, Moving Avg (10) = -197.96
Steps 655360: Avg Reward = -257.89, Moving Avg (10) = -216.44
Steps 671744: Avg Reward = -388.12, Moving Avg (10) = -249.25

Steps 688128: Avg Reward = -397.41, Moving Avg (10) = -259.59
Steps 704512: Avg Reward = -449.84, Moving Avg (10) = -267.56
Steps 720896: Avg Reward = -437.48, Moving Avg (10) = -277.84
Steps 737280: Avg Reward = -128.41, Moving Avg (10) = -273.71
Steps 753664: Avg Reward = -296.03, Moving Avg (10) = -294.19
Steps 770048: Avg Reward = -190.57, Moving Avg (10) = -312.25
Steps 786432: Avg Reward = -156.45, Moving Avg (10) = -293.60
Steps 802816: Avg Reward = -355.59, Moving Avg (10) = -305.78
Steps 819200: Avg Reward = -504.07, Moving Avg (10) = -330.40
Steps 835584: Avg Reward = -761.23, Moving Avg (10) = -367.71
Steps 851968: Avg Reward = -752.61, Moving Avg (10) = -403.23
Steps 868352: Avg Reward = -825.58, Moving Avg (10) = -440.80
Steps 884736: Avg Reward = -798.68, Moving Avg (10) = -476.92
Steps 901120: Avg Reward = -2428.48, Moving Avg (10) = -706.93
Steps 917504: Avg Reward = -1761.60, Moving Avg (10) = -853.49
Steps 933888: Avg Reward = -1023.61, Moving Avg (10) = -936.79
Steps 950272: Avg Reward = -625.85, Moving Avg (10) = -983.73
Steps 966656: Avg Reward = -85.21, Moving Avg (10) = -956.69
Steps 983040: Avg Reward = 1.24, Moving Avg (10) = -906.16
Steps 999424: Avg Reward = -31.20, Moving Avg (10) = -833.16
Steps 1015808: Avg Reward = -62.36, Moving Avg (10) = -764.13



=== GRPO on LunarLander-v2 Final Reward: -100.85

=== Running GRPO_KL on LunarLander-v2 ===

/usr/local/lib/python3.11/dist-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(

/usr/local/lib/python3.11/dist-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(

Steps 16384: Avg Reward = -196.49, Moving Avg (10) = -196.49

Steps 32768: Avg Reward = -133.65, Moving Avg (10) = -165.07

Steps 49152: Avg Reward = -101.56, Moving Avg (10) = -143.90

Steps 65536: Avg Reward = -75.70, Moving Avg (10) = -126.85

Steps 81920: Avg Reward = -66.75, Moving Avg (10) = -114.83

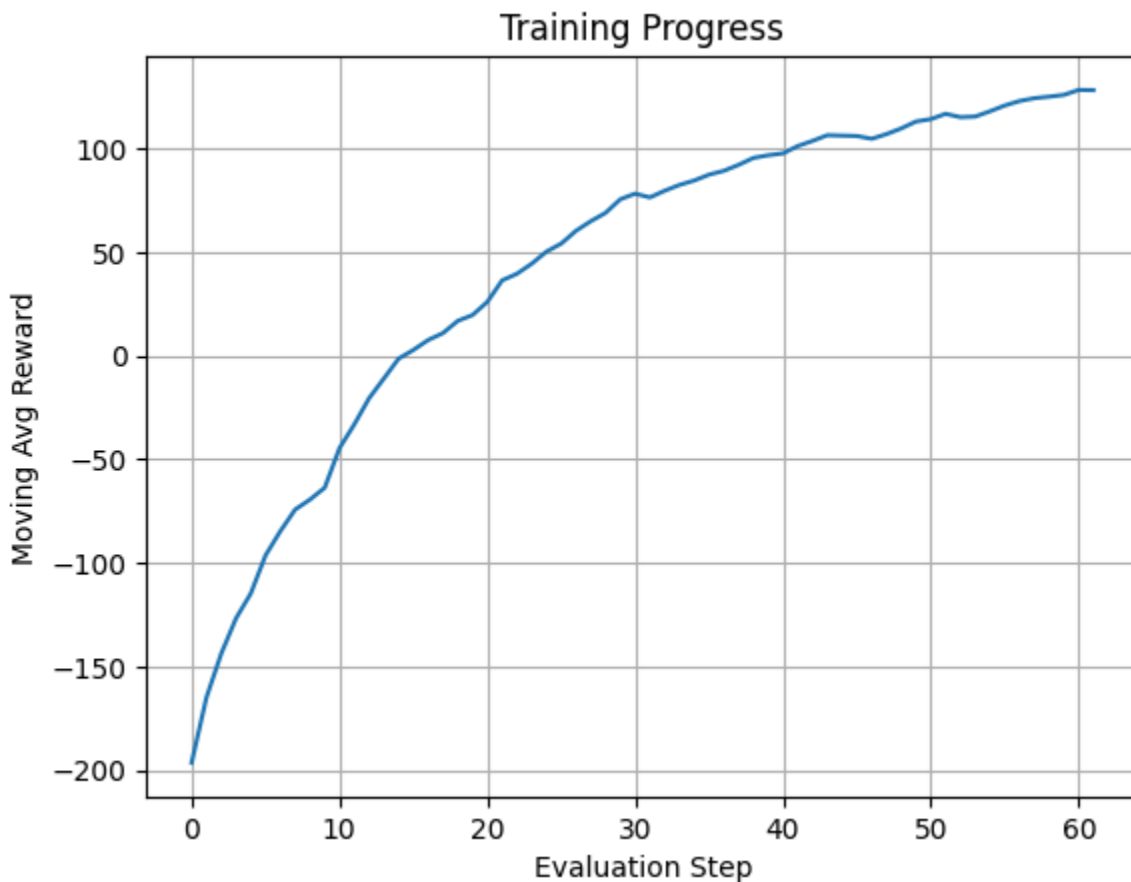
Steps 98304: Avg Reward = -4.37, Moving Avg (10) = -96.42

Steps 114688: Avg Reward = -14.82, Moving Avg (10) = -84.76

Steps 131072: Avg Reward = -1.46, Moving Avg (10) = -74.35

Steps 147456: Avg Reward = -30.94, Moving Avg (10) = -69.53
Steps 163840: Avg Reward = -12.11, Moving Avg (10) = -63.79
Steps 180224: Avg Reward = -5.78, Moving Avg (10) = -44.72
Steps 196608: Avg Reward = -18.97, Moving Avg (10) = -33.25
Steps 212992: Avg Reward = 24.51, Moving Avg (10) = -20.64
Steps 229376: Avg Reward = 19.62, Moving Avg (10) = -11.11
Steps 245760: Avg Reward = 29.36, Moving Avg (10) = -1.50
Steps 262144: Avg Reward = 37.62, Moving Avg (10) = 2.70
Steps 278528: Avg Reward = 33.11, Moving Avg (10) = 7.49
Steps 294912: Avg Reward = 31.93, Moving Avg (10) = 10.83
Steps 311296: Avg Reward = 27.90, Moving Avg (10) = 16.72
Steps 327680: Avg Reward = 16.39, Moving Avg (10) = 19.57
Steps 344064: Avg Reward = 57.20, Moving Avg (10) = 25.87
Steps 360448: Avg Reward = 85.07, Moving Avg (10) = 36.27
Steps 376832: Avg Reward = 56.30, Moving Avg (10) = 39.45
Steps 393216: Avg Reward = 68.29, Moving Avg (10) = 44.32
Steps 409600: Avg Reward = 87.85, Moving Avg (10) = 50.16
Steps 425984: Avg Reward = 76.38, Moving Avg (10) = 54.04
Steps 442368: Avg Reward = 94.84, Moving Avg (10) = 60.21
Steps 458752: Avg Reward = 78.28, Moving Avg (10) = 64.85
Steps 475136: Avg Reward = 68.01, Moving Avg (10) = 68.86
Steps 491520: Avg Reward = 82.32, Moving Avg (10) = 75.45
Steps 507904: Avg Reward = 83.89, Moving Avg (10) = 78.12
Steps 524288: Avg Reward = 67.02, Moving Avg (10) = 76.32
Steps 540672: Avg Reward = 88.59, Moving Avg (10) = 79.55
Steps 557056: Avg Reward = 96.32, Moving Avg (10) = 82.35
Steps 573440: Avg Reward = 109.22, Moving Avg (10) = 84.49
Steps 589824: Avg Reward = 104.70, Moving Avg (10) = 87.32
Steps 606208: Avg Reward = 113.16, Moving Avg (10) = 89.15
Steps 622592: Avg Reward = 107.22, Moving Avg (10) = 92.05
Steps 638976: Avg Reward = 101.39, Moving Avg (10) = 95.38
Steps 655360: Avg Reward = 95.25, Moving Avg (10) = 96.68
Steps 671744: Avg Reward = 92.72, Moving Avg (10) = 97.56
Steps 688128: Avg Reward = 102.66, Moving Avg (10) = 101.12
Steps 704512: Avg Reward = 113.32, Moving Avg (10) = 103.60
Steps 720896: Avg Reward = 123.83, Moving Avg (10) = 106.35
Steps 737280: Avg Reward = 107.20, Moving Avg (10) = 106.15
Steps 753664: Avg Reward = 102.82, Moving Avg (10) = 105.96
Steps 770048: Avg Reward = 99.84, Moving Avg (10) = 104.63
Steps 786432: Avg Reward = 129.71, Moving Avg (10) = 106.87
Steps 802816: Avg Reward = 128.88, Moving Avg (10) = 109.62
Steps 819200: Avg Reward = 128.63, Moving Avg (10) = 112.96
Steps 835584: Avg Reward = 104.16, Moving Avg (10) = 114.10
Steps 851968: Avg Reward = 128.30, Moving Avg (10) = 116.67

Steps 868352: Avg Reward = 97.14, Moving Avg (10) = 115.05
Steps 884736: Avg Reward = 126.76, Moving Avg (10) = 115.34
Steps 901120: Avg Reward = 132.57, Moving Avg (10) = 117.88
Steps 917504: Avg Reward = 130.69, Moving Avg (10) = 120.67
Steps 933888: Avg Reward = 121.52, Moving Avg (10) = 122.84
Steps 950272: Avg Reward = 143.41, Moving Avg (10) = 124.21
Steps 966656: Avg Reward = 136.18, Moving Avg (10) = 124.94
Steps 983040: Avg Reward = 137.29, Moving Avg (10) = 125.80
Steps 999424: Avg Reward = 128.00, Moving Avg (10) = 128.19
Steps 1015808: Avg Reward = 127.36, Moving Avg (10) = 128.09



=== GRPO_KL on LunarLander-v2 Final Reward: 127.29

=== Running PPO_CLIP on LunarLander-v2 ===

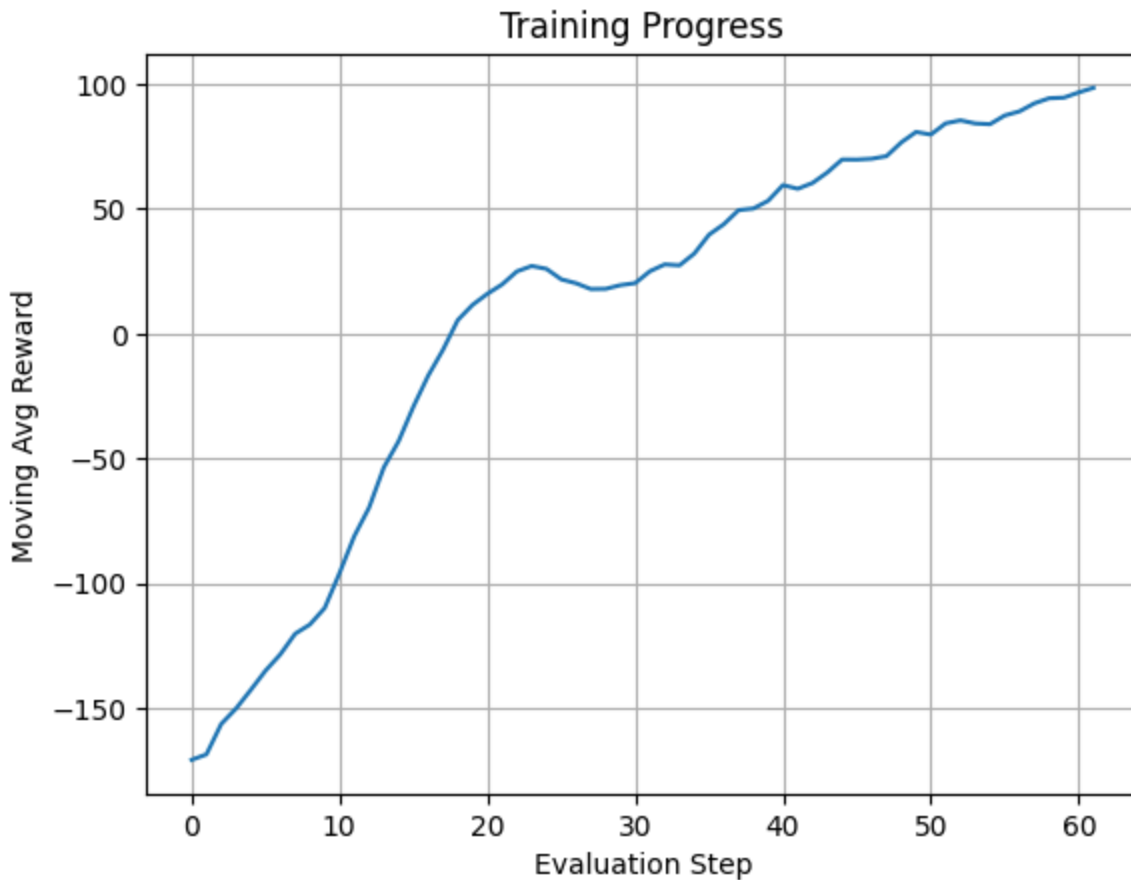
```
/usr/local/lib/python3.11/dist-packages/gym/core.py:317: DeprecationWarning: WARN:
Initializing wrapper in old step API which returns one bool instead of two. It is recommended to
set `new_step_api=True` to use new step API. This will be the default behaviour in future.
  deprecation(
/usr/local/lib/python3.11/dist-packages/gym/wrappers/step_api_compatibility.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(

Steps 16384: Avg Reward = -170.59, Moving Avg (10) = -170.59
Steps 32768: Avg Reward = -166.24, Moving Avg (10) = -168.41
Steps 49152: Avg Reward = -131.88, Moving Avg (10) = -156.24
Steps 65536: Avg Reward = -131.61, Moving Avg (10) = -150.08
Steps 81920: Avg Reward = -112.67, Moving Avg (10) = -142.60
Steps 98304: Avg Reward = -96.33, Moving Avg (10) = -134.89
Steps 114688: Avg Reward = -89.25, Moving Avg (10) = -128.37
Steps 131072: Avg Reward = -62.45, Moving Avg (10) = -120.13
Steps 147456: Avg Reward = -86.77, Moving Avg (10) = -116.42
Steps 163840: Avg Reward = -50.32, Moving Avg (10) = -109.81
Steps 180224: Avg Reward = -31.48, Moving Avg (10) = -95.90
Steps 196608: Avg Reward = -17.78, Moving Avg (10) = -81.05
Steps 212992: Avg Reward = -17.80, Moving Avg (10) = -69.65
Steps 229376: Avg Reward = 29.57, Moving Avg (10) = -53.53
Steps 245760: Avg Reward = -7.67, Moving Avg (10) = -43.03
Steps 262144: Avg Reward = 41.87, Moving Avg (10) = -29.21
Steps 278528: Avg Reward = 34.68, Moving Avg (10) = -16.81
Steps 294912: Avg Reward = 41.07, Moving Avg (10) = -6.46
Steps 311296: Avg Reward = 32.16, Moving Avg (10) = 5.43
Steps 327680: Avg Reward = 10.61, Moving Avg (10) = 11.52
Steps 344064: Avg Reward = 12.76, Moving Avg (10) = 15.95
Steps 360448: Avg Reward = 19.85, Moving Avg (10) = 19.71
Steps 376832: Avg Reward = 34.40, Moving Avg (10) = 24.93
Steps 393216: Avg Reward = 51.17, Moving Avg (10) = 27.09
Steps 409600: Avg Reward = -19.05, Moving Avg (10) = 25.95
Steps 425984: Avg Reward = -0.09, Moving Avg (10) = 21.76
Steps 442368: Avg Reward = 19.04, Moving Avg (10) = 20.19
Steps 458752: Avg Reward = 17.79, Moving Avg (10) = 17.86
Steps 475136: Avg Reward = 32.90, Moving Avg (10) = 17.94
Steps 491520: Avg Reward = 25.45, Moving Avg (10) = 19.42
Steps 507904: Avg Reward = 20.80, Moving Avg (10) = 20.23
Steps 524288: Avg Reward = 67.66, Moving Avg (10) = 25.01
Steps 540672: Avg Reward = 61.65, Moving Avg (10) = 27.73
Steps 557056: Avg Reward = 47.37, Moving Avg (10) = 27.35
Steps 573440: Avg Reward = 28.19, Moving Avg (10) = 32.08
Steps 589824: Avg Reward = 75.61, Moving Avg (10) = 39.65
Steps 606208: Avg Reward = 60.86, Moving Avg (10) = 43.83
Steps 622592: Avg Reward = 74.19, Moving Avg (10) = 49.47
Steps 638976: Avg Reward = 39.29, Moving Avg (10) = 50.11
Steps 655360: Avg Reward = 56.57, Moving Avg (10) = 53.22
Steps 671744: Avg Reward = 83.44, Moving Avg (10) = 59.48

Steps 688128: Avg Reward = 52.85, Moving Avg (10) = 58.00
Steps 704512: Avg Reward = 84.96, Moving Avg (10) = 60.33
Steps 720896: Avg Reward = 89.55, Moving Avg (10) = 64.55
Steps 737280: Avg Reward = 79.95, Moving Avg (10) = 69.73
Steps 753664: Avg Reward = 75.22, Moving Avg (10) = 69.69
Steps 770048: Avg Reward = 64.15, Moving Avg (10) = 70.02
Steps 786432: Avg Reward = 84.62, Moving Avg (10) = 71.06
Steps 802816: Avg Reward = 94.29, Moving Avg (10) = 76.56
Steps 819200: Avg Reward = 98.53, Moving Avg (10) = 80.76
Steps 835584: Avg Reward = 72.86, Moving Avg (10) = 79.70
Steps 851968: Avg Reward = 97.74, Moving Avg (10) = 84.19
Steps 868352: Avg Reward = 97.07, Moving Avg (10) = 85.40
Steps 884736: Avg Reward = 77.06, Moving Avg (10) = 84.15
Steps 901120: Avg Reward = 77.10, Moving Avg (10) = 83.87
Steps 917504: Avg Reward = 110.01, Moving Avg (10) = 87.34
Steps 933888: Avg Reward = 80.72, Moving Avg (10) = 89.00
Steps 950272: Avg Reward = 116.57, Moving Avg (10) = 92.20
Steps 966656: Avg Reward = 114.77, Moving Avg (10) = 94.24
Steps 983040: Avg Reward = 100.96, Moving Avg (10) = 94.49
Steps 999424: Avg Reward = 93.67, Moving Avg (10) = 96.57
Steps 1015808: Avg Reward = 116.20, Moving Avg (10) = 98.41



=== PPO_CLIP on LunarLander-v2 Final Reward: 104.27

=== Running PPO_KL on LunarLander-v2 ===

/usr/local/lib/python3.11/dist-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(

/usr/local/lib/python3.11/dist-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(

Steps 16384: Avg Reward = -183.52, Moving Avg (10) = -183.52

Steps 32768: Avg Reward = -122.54, Moving Avg (10) = -153.03

Steps 49152: Avg Reward = -113.24, Moving Avg (10) = -139.77

Steps 65536: Avg Reward = -73.70, Moving Avg (10) = -123.25

Steps 81920: Avg Reward = -62.08, Moving Avg (10) = -111.02

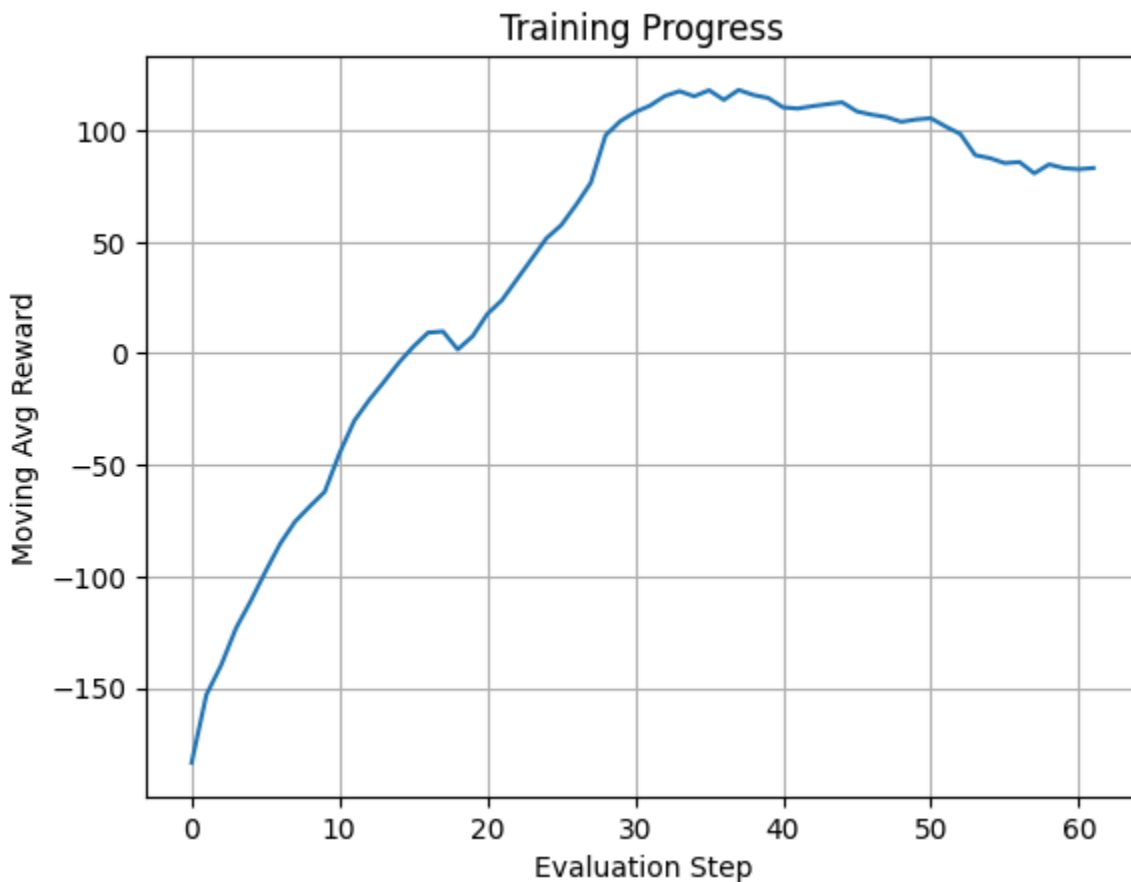
Steps 98304: Avg Reward = -31.50, Moving Avg (10) = -97.76

Steps 114688: Avg Reward = -9.96, Moving Avg (10) = -85.22

Steps 131072: Avg Reward = -6.61, Moving Avg (10) = -75.39

Steps 147456: Avg Reward = -13.95, Moving Avg (10) = -68.57
Steps 163840: Avg Reward = -3.70, Moving Avg (10) = -62.08
Steps 180224: Avg Reward = -11.43, Moving Avg (10) = -44.87
Steps 196608: Avg Reward = 23.78, Moving Avg (10) = -30.24
Steps 212992: Avg Reward = -20.83, Moving Avg (10) = -21.00
Steps 229376: Avg Reward = 7.88, Moving Avg (10) = -12.84
Steps 245760: Avg Reward = 23.37, Moving Avg (10) = -4.30
Steps 262144: Avg Reward = 42.02, Moving Avg (10) = 3.06
Steps 278528: Avg Reward = 52.65, Moving Avg (10) = 9.32
Steps 294912: Avg Reward = -1.45, Moving Avg (10) = 9.83
Steps 311296: Avg Reward = -94.68, Moving Avg (10) = 1.76
Steps 327680: Avg Reward = 55.03, Moving Avg (10) = 7.63
Steps 344064: Avg Reward = 88.63, Moving Avg (10) = 17.64
Steps 360448: Avg Reward = 86.88, Moving Avg (10) = 23.95
Steps 376832: Avg Reward = 71.27, Moving Avg (10) = 33.16
Steps 393216: Avg Reward = 99.48, Moving Avg (10) = 42.32
Steps 409600: Avg Reward = 116.28, Moving Avg (10) = 51.61
Steps 425984: Avg Reward = 101.03, Moving Avg (10) = 57.51
Steps 442368: Avg Reward = 142.94, Moving Avg (10) = 66.54
Steps 458752: Avg Reward = 97.95, Moving Avg (10) = 76.48
Steps 475136: Avg Reward = 118.32, Moving Avg (10) = 97.78
Steps 491520: Avg Reward = 118.56, Moving Avg (10) = 104.13
Steps 507904: Avg Reward = 129.15, Moving Avg (10) = 108.19
Steps 524288: Avg Reward = 115.62, Moving Avg (10) = 111.06
Steps 540672: Avg Reward = 114.26, Moving Avg (10) = 115.36
Steps 557056: Avg Reward = 121.38, Moving Avg (10) = 117.55
Steps 573440: Avg Reward = 93.49, Moving Avg (10) = 115.27
Steps 589824: Avg Reward = 128.23, Moving Avg (10) = 117.99
Steps 606208: Avg Reward = 99.72, Moving Avg (10) = 113.67
Steps 622592: Avg Reward = 142.41, Moving Avg (10) = 118.11
Steps 638976: Avg Reward = 95.72, Moving Avg (10) = 115.85
Steps 655360: Avg Reward = 105.03, Moving Avg (10) = 114.50
Steps 671744: Avg Reward = 87.31, Moving Avg (10) = 110.32
Steps 688128: Avg Reward = 110.26, Moving Avg (10) = 109.78
Steps 704512: Avg Reward = 124.92, Moving Avg (10) = 110.85
Steps 720896: Avg Reward = 130.41, Moving Avg (10) = 111.75
Steps 737280: Avg Reward = 101.72, Moving Avg (10) = 112.57
Steps 753664: Avg Reward = 87.54, Moving Avg (10) = 108.50
Steps 770048: Avg Reward = 85.04, Moving Avg (10) = 107.03
Steps 786432: Avg Reward = 131.74, Moving Avg (10) = 105.97
Steps 802816: Avg Reward = 74.22, Moving Avg (10) = 103.82
Steps 819200: Avg Reward = 115.46, Moving Avg (10) = 104.86
Steps 835584: Avg Reward = 92.95, Moving Avg (10) = 105.43
Steps 851968: Avg Reward = 73.28, Moving Avg (10) = 101.73

Steps 868352: Avg Reward = 91.28, Moving Avg (10) = 98.36
 Steps 884736: Avg Reward = 35.68, Moving Avg (10) = 88.89
 Steps 901120: Avg Reward = 88.20, Moving Avg (10) = 87.54
 Steps 917504: Avg Reward = 65.28, Moving Avg (10) = 85.31
 Steps 933888: Avg Reward = 89.86, Moving Avg (10) = 85.79
 Steps 950272: Avg Reward = 80.97, Moving Avg (10) = 80.72
 Steps 966656: Avg Reward = 114.46, Moving Avg (10) = 84.74
 Steps 983040: Avg Reward = 98.50, Moving Avg (10) = 83.05
 Steps 999424: Avg Reward = 88.28, Moving Avg (10) = 82.58
 Steps 1015808: Avg Reward = 78.58, Moving Avg (10) = 83.11



=== PPO_KL on LunarLander-v2 Final Reward: 105.24

2m4313s

=== Running A2C on LunarLander-v2 ===

/usr/local/lib/python3.11/dist-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(

/usr/local/lib/python3.11/dist-packages/gym/wrappers/step_api_compatibility.py:39:

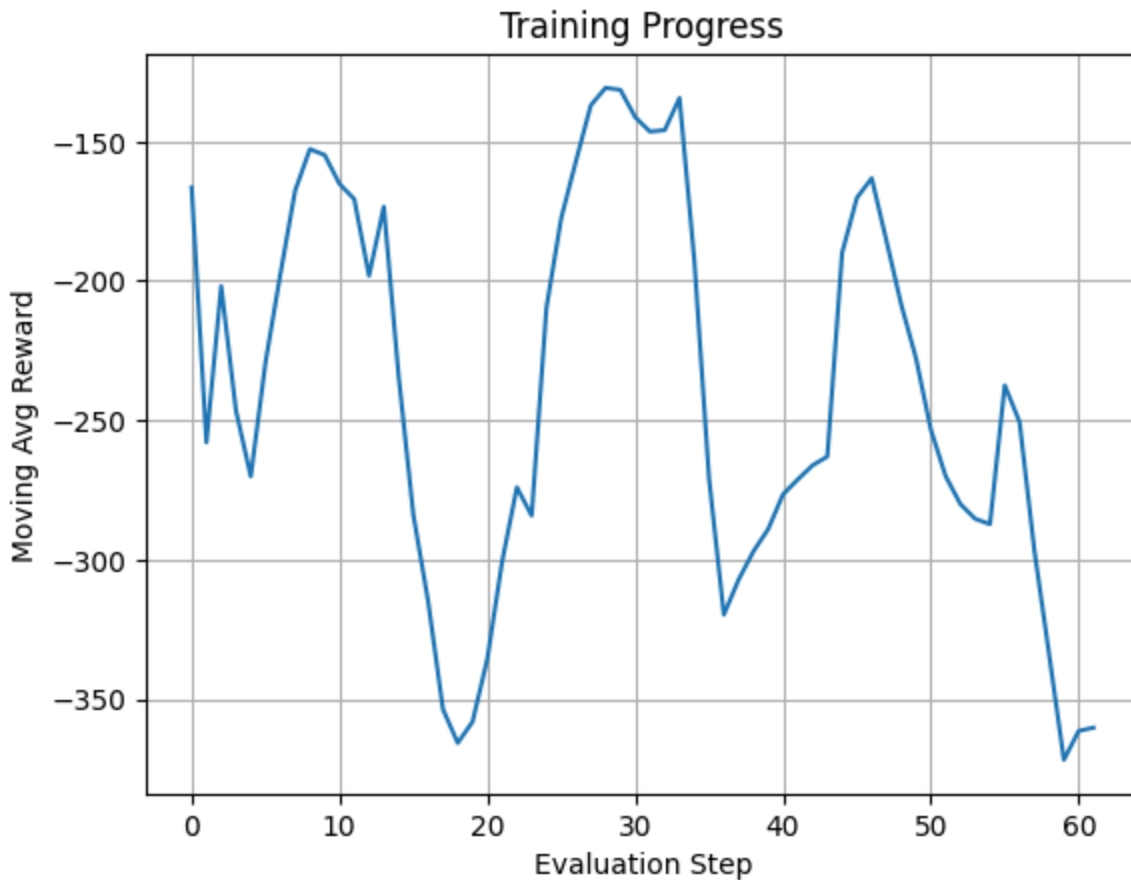
DeprecationWarning: WARN: Initializing environment in old step API which returns one bool

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(

Steps 16384: Avg Reward = -166.44, Moving Avg (10) = -166.44
Steps 32768: Avg Reward = -349.38, Moving Avg (10) = -257.91
Steps 49152: Avg Reward = -89.75, Moving Avg (10) = -201.86
Steps 65536: Avg Reward = -380.54, Moving Avg (10) = -246.53
Steps 81920: Avg Reward = -364.53, Moving Avg (10) = -270.13
Steps 98304: Avg Reward = -23.80, Moving Avg (10) = -229.07
Steps 114688: Avg Reward = -8.88, Moving Avg (10) = -197.62
Steps 131072: Avg Reward = 42.48, Moving Avg (10) = -167.61
Steps 147456: Avg Reward = -33.28, Moving Avg (10) = -152.68
Steps 163840: Avg Reward = -173.96, Moving Avg (10) = -154.81
Steps 180224: Avg Reward = -269.91, Moving Avg (10) = -165.15
Steps 196608: Avg Reward = -404.17, Moving Avg (10) = -170.63
Steps 212992: Avg Reward = -364.79, Moving Avg (10) = -198.14
Steps 229376: Avg Reward = -132.81, Moving Avg (10) = -173.37
Steps 245760: Avg Reward = -970.74, Moving Avg (10) = -233.99
Steps 262144: Avg Reward = -523.49, Moving Avg (10) = -283.96
Steps 278528: Avg Reward = -319.62, Moving Avg (10) = -315.03
Steps 294912: Avg Reward = -343.89, Moving Avg (10) = -353.67
Steps 311296: Avg Reward = -154.08, Moving Avg (10) = -365.74
Steps 327680: Avg Reward = -98.02, Moving Avg (10) = -358.15
Steps 344064: Avg Reward = -44.52, Moving Avg (10) = -335.61
Steps 360448: Avg Reward = -55.41, Moving Avg (10) = -300.74
Steps 376832: Avg Reward = -97.77, Moving Avg (10) = -274.03
Steps 393216: Avg Reward = -234.22, Moving Avg (10) = -284.17
Steps 409600: Avg Reward = -224.72, Moving Avg (10) = -209.57
Steps 425984: Avg Reward = -202.06, Moving Avg (10) = -177.43
Steps 442368: Avg Reward = -114.30, Moving Avg (10) = -156.90
Steps 458752: Avg Reward = -144.23, Moving Avg (10) = -136.93
Steps 475136: Avg Reward = -91.71, Moving Avg (10) = -130.70
Steps 491520: Avg Reward = -105.18, Moving Avg (10) = -131.41
Steps 507904: Avg Reward = -142.38, Moving Avg (10) = -141.20
Steps 524288: Avg Reward = -107.31, Moving Avg (10) = -146.39
Steps 540672: Avg Reward = -92.64, Moving Avg (10) = -145.88
Steps 557056: Avg Reward = -118.78, Moving Avg (10) = -134.33
Steps 573440: Avg Reward = -804.99, Moving Avg (10) = -192.36
Steps 589824: Avg Reward = -987.64, Moving Avg (10) = -270.92
Steps 606208: Avg Reward = -602.43, Moving Avg (10) = -319.73
Steps 622592: Avg Reward = -18.37, Moving Avg (10) = -307.14
Steps 638976: Avg Reward = 9.55, Moving Avg (10) = -297.02
Steps 655360: Avg Reward = -25.21, Moving Avg (10) = -289.02
Steps 671744: Avg Reward = -17.86, Moving Avg (10) = -276.57

Steps 688128: Avg Reward = -53.72, Moving Avg (10) = -271.21
Steps 704512: Avg Reward = -42.12, Moving Avg (10) = -266.16
Steps 720896: Avg Reward = -86.86, Moving Avg (10) = -262.97
Steps 737280: Avg Reward = -74.05, Moving Avg (10) = -189.87
Steps 753664: Avg Reward = -790.46, Moving Avg (10) = -170.15
Steps 770048: Avg Reward = -532.11, Moving Avg (10) = -163.12
Steps 786432: Avg Reward = -243.27, Moving Avg (10) = -185.61
Steps 802816: Avg Reward = -219.32, Moving Avg (10) = -208.50
Steps 819200: Avg Reward = -218.05, Moving Avg (10) = -227.78
Steps 835584: Avg Reward = -273.03, Moving Avg (10) = -253.30
Steps 851968: Avg Reward = -220.85, Moving Avg (10) = -270.01
Steps 868352: Avg Reward = -142.59, Moving Avg (10) = -280.06
Steps 884736: Avg Reward = -139.44, Moving Avg (10) = -285.32
Steps 901120: Avg Reward = -93.51, Moving Avg (10) = -287.26
Steps 917504: Avg Reward = -292.28, Moving Avg (10) = -237.45
Steps 933888: Avg Reward = -663.03, Moving Avg (10) = -250.54
Steps 950272: Avg Reward = -702.60, Moving Avg (10) = -296.47
Steps 966656: Avg Reward = -593.85, Moving Avg (10) = -333.92
Steps 983040: Avg Reward = -596.90, Moving Avg (10) = -371.81
Steps 999424: Avg Reward = -169.47, Moving Avg (10) = -361.45
Steps 1015808: Avg Reward = -208.92, Moving Avg (10) = -360.26



=== A2C on LunarLander-v2 Final Reward: -168.62

3h8m32s

=== Running GRPO on MountainCar-v0 ===

/usr/local/lib/python3.11/dist-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(

/usr/local/lib/python3.11/dist-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(