```
!pip install swig
```

```
Collecting swig
  Downloading swig-4.4.0-py3-none-manylinux_2_12_x86_64.manylinux2010_x86_64.whl.metadata (3.5 kB)
Downloading swig-4.4.0-py3-none-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (1.9 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.9/1.9 MB 81.8 MB/s eta 0:00:00
Installing collected packages: swig
Successfully installed swig-4.4.0
```

```
!pip install stable_baselines3 gymnasium[box2d]
```

```
Collecting stable_baselines3
  Downloading stable_baselines3-2.7.0-py3-none-any.whl.metadata (4.8 kB)
Requirement already satisfied: gymnasium[box2d] in /usr/local/lib/python3.12/dist-packages (1.2.2)
Requirement already satisfied: numpy<3.0,>=1.20 in /usr/local/lib/python3.12/dist-packages (from stable_baselines3) (2.0.2)
Requirement already satisfied: torch<3.0,>=2.3 in /usr/local/lib/python3.12/dist-packages (from stable_baselines3) (2.9.0+cu126)
Requirement already satisfied: cloudpickle in /usr/local/lib/python3.12/dist-packages (from stable_baselines3) (3.1.2)
Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (from stable_baselines3) (2.2.2)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.12/dist-packages (from stable_baselines3) (3.10.0)
Requirement already satisfied: typing-extensions>=4.3.0 in /usr/local/lib/python3.12/dist-packages (from gymnasium[box2d]) (4.15.0)
Requirement already satisfied: farama-notifications>=0.0.1 in /usr/local/lib/python3.12/dist-packages (from gymnasium[box2d]) (0.0.4)
Collecting box2d-py==2.3.5 (from gymnasium[box2d])
  Downloading box2d-py-2.3.5.tar.gz (374 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 374.4/374.4 kB 29.1 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: pygame>=2.1.3 in /usr/local/lib/python3.12/dist-packages (from gymnasium[box2d]) (2.6.1)
Requirement already satisfied: swig==4.* in /usr/local/lib/python3.12/dist-packages (from gymnasium[box2d]) (4.4.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (3.20.0)
Requirement already satisfied: setuptools in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (75.2.0)
Requirement already satisfied: sympy>=1.13.3 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (1.13.3)
Requirement already satisfied: networkx>=2.5.1 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (3.5)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (3.1.6)
Requirement already satisfied: fsspec>=0.8.5 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (2025.3.0)
Requirement already satisfied: nvidia-cuda-nvrtc-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (12.6.77)
Requirement already satisfied: nvidia-cuda-runtime-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (12.6.77
Requirement already satisfied: nvidia-cuda-cupti-cu12==12.6.80 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (12.6.80)
Requirement already satisfied: nvidia-cudnn-cu12==9.10.2.21 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (9.10.2.21)
Requirement already satisfied: nvidia-cublas-cu12==12.6.4.1 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (12.6.4.1)
Requirement already satisfied: nvidia-cufft-cu12==11.3.0.4 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (11.3.0.4)
Requirement already satisfied: nvidia-curand-cu12==10.3.7.77 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (10.3.7.77)
Requirement already satisfied: nvidia-cusolver-cu12==11.7.1.2 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (11.7.1.2)
Requirement already satisfied: nvidia-cusparse-cu12==12.5.4.2 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (12.5.4.2)
Requirement already satisfied: nvidia-cusparselt-cu12==0.7.1 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (0.7.1)
Requirement already satisfied: nvidia-nccl-cu12==2.27.5 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (2.27.5)
Requirement already satisfied: nvidia-nvshmem-cu12==3.3.20 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (3.3.20)
Requirement already satisfied: nvidia-nvtx-cu12==12.6.77 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (12.6.77)
Requirement already satisfied: nvidia-nvjitlink-cu12==12.6.85 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (12.6.85)
Requirement already satisfied: nvidia-cufile-cu12==1.11.1.6 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (1.11.1.6)
Requirement already satisfied: triton==3.5.0 in /usr/local/lib/python3.12/dist-packages (from torch<3.0,>=2.3->stable_baselines3) (3.5.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->stable_baselines3) (1.3.3)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.12/dist-packages (from matplotlib->stable_baselines3) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib->stable_baselines3) (4.60.1)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->stable_baselines3) (1.4.9)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.12/dist-packages (from matplotlib->stable_baselines3) (25.0)
Requirement already satisfied: pillow>=8 in /usr/local/lib/python3.12/dist-packages (from matplotlib->stable_baselines3) (11.3.0)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.12/dist-packages (from matplotlib->stable_baselines3) (3.2.5)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.12/dist-packages (from matplotlib->stable_baselines3) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas->stable_baselines3) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas->stable_baselines3) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.7->matplotlib->stable_baselines3) (1.17.0)
```

```
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3->torch<3.0,>=2.3->stable_baselines3) (1.3.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.12/dist-packages (from jinja2->torch<3.0,>=2.3->stable_baselines3) (3.0.3)
Downloading stable_baselines3-2.7.0-py3-none-any.whl (187 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 187.2/187.2 kB 18.3 MB/s eta 0:00:00
Building wheels for collected packages: box2d-py
  Building wheel for box2d-py (setup.py) ... done
  Created wheel for box2d-py: filename=box2d_py-2.3.5-cp312-cp312-linux_x86_64.whl size=2398998 sha256=959307cfc226c6163272eb02fed9697ca18371b77f2a39c0b4fe2bf
  Stored in directory: /root/.cache/pip/wheels/2a/e9/60/774da0bcd07f7dc7761a8590fa2d065e4069568e78dcdc3318
```

```
import os, platform, warnings
warnings.filterwarnings("ignore", category=DeprecationWarning)
warnings.filterwarnings("ignore", message="This system does not have CUDA")
warnings.filterwarnings("ignore", message="pkg_resources is deprecated")

import numpy as np
import torch
import matplotlib.pyplot as plt

import gymnasium as gym
from stable_baselines3 import SAC
from stable_baselines3.common.atari_wrappers import WarpFrame  # 84x84 grayscale
from stable_baselines3.common.env_util import make_vec_env
from stable_baselines3.common.vec_env import (
    VecFrameStack, VecTransposeImage, VecNormalize, VecVideoRecorder
)
from stable_baselines3.common.callbacks import EvalCallback
from stable_baselines3.common.evaluation import evaluate_policy
from stable_baselines3.common.torch_layers import NatureCNN
from importlib.metadata import version

print(f"Python: {platform.python_version()}")
print(f"Torch:  {version('torch')} | CUDA avail: {torch.cuda.is_available()} | torch.cuda: {torch.version.cuda}")
print(f"Gymnasium: {version('gymnasium')}")
print(f"SB3: {version('stable_baselines3')}")
print(f"Numpy: {version('numpy')}  Matplotlib: {version('matplotlib')}")
```

```
Python: 3.12.12
Torch:  2.9.0+cu126 | CUDA avail: True | torch.cuda: 12.6
Gymnasium: 1.2.2
SB3: 2.7.0
Numpy: 2.0.2  Matplotlib: 3.10.0
Gym has been unmaintained since 2022 and does not support NumPy 2.0 amongst other critical functionality.
Please upgrade to Gymnasium, the maintained drop-in replacement of Gym, or contact the authors of your software and request that they upgrade.
See the migration guide at https://gymnasium.farama.org/introduction/migration_guide/ for additional information.
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated and scheduled for removal ir
  return datetime.utcnow().replace(tzinfo=utc)
```

```
# ----------------------
# Config (faster run)
# ----------------------
ENV_ID = "CarRacing-v3"
LOG_DIR = f"./logs/{ENV_ID}"
VIDEO_DIR = "./videos"
os.makedirs(LOG_DIR, exist_ok=True)
os.makedirs(VIDEO_DIR, exist_ok=True)

TOTAL_STEPS = 500_000       # ↓ from 1M
```

```
EVAL_FREQ = 20_000          # evaluate a bit less often
EVAL_EPISODES = 5           # fewer episodes during training to save time
FINAL_EVAL_EPISODES = 20    # do a fuller eval at the end for your report
SEED = 42

USE_GRAY = True
WRAPPER_CLASS = WarpFrame if USE_GRAY else None

def make_training_env(seed=SEED):
    env = make_vec_env(ENV_ID, n_envs=1, seed=seed, wrapper_class=WRAPPER_CLASS)
    env = VecFrameStack(env, n_stack=4)
    env = VecTransposeImage(env)
    # Reward normalization helps SAC stability on pixels
    env = VecNormalize(env, norm_obs=False, norm_reward=True, clip_reward=10.0)
    return env

def make_eval_env(seed=SEED+1):
    env = make_vec_env(ENV_ID, n_envs=1, seed=seed, wrapper_class=WRAPPER_CLASS)
    env = VecFrameStack(env, n_stack=4)
    env = VecTransposeImage(env)
    env = VecNormalize(env, norm_obs=False, norm_reward=True, clip_reward=10.0, training=False)
    return env

train_env = make_training_env(SEED)
eval_env = make_eval_env(SEED + 1)

policy_kwargs = dict(
    features_extractor_class=NatureCNN,
    features_extractor_kwargs=dict(features_dim=512),  # stronger CNN head
    net_arch=[256, 256],
)

eval_callback = EvalCallback(
    eval_env,
    best_model_save_path=LOG_DIR,
    log_path=LOG_DIR,
    eval_freq=EVAL_FREQ,
    n_eval_episodes=EVAL_EPISODES,
    deterministic=True,
    render=False,
)
```

```
/usr/local/lib/python3.12/dist-packages/pygame/pkgdata.py:25: DeprecationWarning: pkg_resources is deprecated as an API. See https://setuptools.pypa.io/en/lates
  from pkg_resources import resource_stream, resource_exists
/usr/local/lib/python3.12/dist-packages/pkg_resources/__init__.py:3154: DeprecationWarning: Deprecated call to `pkg_resources.declare_namespace('google')`.
Implementing implicit namespace packages (as specified in PEP 420) is preferred to `pkg_resources.declare_namespace`. See https://setuptools.pypa.io/en/latest/r
  declare_namespace(pkg)
/usr/local/lib/python3.12/dist-packages/pkg_resources/__init__.py:3154: DeprecationWarning: Deprecated call to `pkg_resources.declare_namespace('google.cloud')`
Implementing implicit namespace packages (as specified in PEP 420) is preferred to `pkg_resources.declare_namespace`. See https://setuptools.pypa.io/en/latest/r
  declare_namespace(pkg)
/usr/local/lib/python3.12/dist-packages/pkg_resources/__init__.py:3154: DeprecationWarning: Deprecated call to `pkg_resources.declare_namespace('sphinxcontrib')`
Implementing implicit namespace packages (as specified in PEP 420) is preferred to `pkg_resources.declare_namespace`. See https://setuptools.pypa.io/en/latest/r
  declare_namespace(pkg)
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated and scheduled for removal in
  return datetime.utcnow().replace(tzinfo=utc)
```

```
# --------------------
# SAC hyperparams (slightly lighter/faster)
# --------------------
model = SAC(
    "CnnPolicy",
    train_env,
    learning_rate=1e-4,
    buffer_size=500_000,        # ↓ from 1,000,000 to reduce RAM/time
    batch_size=256,
    learning_starts=5_000,      # ↓ warmup so learning begins earlier
    gamma=0.99,
    tau=0.005,
    ent_coef="auto",
    target_entropy="auto",
    train_freq=(1, "step"),
    gradient_steps=1,
    policy_kwargs=policy_kwargs,
    verbose=1,
    tensorboard_log=LOG_DIR,
    seed=SEED,
)
```

```
Using cuda device
/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated and scheduled for removal ir
  return datetime.utcnow().replace(tzinfo=utc)
```

```
# --------------------
# Train
# --------------------
model.learn(total_timesteps=TOTAL_STEPS, callback=eval_callback, progress_bar=True)

# Save final model and normalization stats
final_model_path = os.path.join(LOG_DIR, "sac_car_racing_final_500k")
model.save(final_model_path)
train_env.save(os.path.join(LOG_DIR, "vecnormalize.pkl"))
```

```
Logging to ./logs/CarRacing-v3/SAC_1
 100% ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 500,000/500,000  [ 7:26:35 < 0:00:00 , ? it/s ]
/usr/local/lib/python3.12/dist-packages/ipywidgets/widgets/widget_output.py:111: DeprecationWarning:
Kernel._parent_header is deprecated in ipykernel 6. Use .get_parent()
  if ip and hasattr(ip, 'kernel') and hasattr(ip.kernel, '_parent_header'):
---------------------------------
| rollout/            |         |
|    ep_len_mean      | 1e+03   |
|    ep_rew_mean      | -31.6   |
| time/               |         |
|    episodes         | 4       |
|    fps              | 94      |
|    time_elapsed     | 42      |
|    total_timesteps  | 4000    |
---------------------------------
---------------------------------
| rollout/            |         |
|    ep_len_mean      | 1e+03   |
|    ep_rew_mean      | -35     |
| time/               |         |
|    episodes         | 8       |
|    fps              | 38      |
|    time_elapsed     | 206     |
|    total_timesteps  | 8000    |
| train/              |         |
|    actor_loss       | -23.3   |
|    critic_loss      | 0.0994  |
|    ent_coef         | 0.741   |
|    ent_coef_loss    | -1.51   |
|    learning_rate    | 0.0001  |
|    n_updates        | 2999    |
---------------------------------
---------------------------------
| rollout/            |         |
|    ep_len_mean      | 1e+03   |
|    ep_rew_mean      | -34.5   |
| time/               |         |
|    episodes         | 12      |
|    fps              | 29      |
|    time_elapsed     | 406     |
|    total_timesteps  | 12000   |
| train/              |         |
|    actor_loss       | -40.1   |
|    critic_loss      | 0.107   |
|    ent_coef         | 0.497   |
|    ent_coef_loss    | -3.53   |
|    learning_rate    | 0.0001  |
|    n_updates        | 6999    |
---------------------------------
---------------------------------
| rollout/            |         |
|    ep_len_mean      | 1e+03   |
|    ep_rew_mean      | -33.6   |
| time/               |         |
|    episodes         | 16      |
|    fps              | 26      |
|    time_elapsed     | 606     |
|    total_timesteps  | 16000   |
| train/              |         |
|    actor_loss       | -47.3   |
|    critic_loss      | 0.0478  |
|    ent_coef         | 0.333   |
|    ent_coef_loss    | -5.57   |
|    learning_rate    | 0.0001  |
```

```
|    n_updates       | 10999    |
----------------------------------
Eval num_timesteps=20000, episode_reward=-18.57 +/- 16.33
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -18.6    |
| time/              |          |
|    total_timesteps | 20000    |
| train/             |          |
|    actor_loss      | -48      |
|    critic_loss     | 0.0523   |
|    ent_coef        | 0.223    |
|    ent_coef_loss   | -7.57    |
|    learning_rate   | 0.0001   |
|    n_updates       | 14999    |
----------------------------------
New best mean reward!
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -32.9    |
| time/              |          |
|    episodes        | 20       |
|    fps             | 23       |
|    time_elapsed    | 867      |
|    total_timesteps | 20000    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -32.5    |
| time/              |          |
|    episodes        | 24       |
|    fps             | 22       |
|    time_elapsed    | 1070     |
|    total_timesteps | 24000    |
| train/             |          |
|    actor_loss      | -45.5    |
|    critic_loss     | 0.0299   |
|    ent_coef        | 0.15     |
|    ent_coef_loss   | -9.62    |
|    learning_rate   | 0.0001   |
|    n_updates       | 18999    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -32.9    |
| time/              |          |
|    episodes        | 28       |
|    fps             | 21       |
|    time_elapsed    | 1273     |
|    total_timesteps | 28000    |
| train/             |          |
|    actor_loss      | -41.2    |
|    critic_loss     | 0.0263   |
|    ent_coef        | 0.1      |
|    ent_coef_loss   | -11.6    |
|    learning_rate   | 0.0001   |
|    n_updates       | 22999    |
----------------------------------
----------------------------------
```

```
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -32.9    |
| time/              |          |
|    episodes        | 32       |
|    fps             | 21       |
|    time_elapsed    | 1476     |
|    total_timesteps | 32000    |
| train/             |          |
|    actor_loss      | -36.4    |
|    critic_loss     | 0.059    |
|    ent_coef        | 0.0672   |
|    ent_coef_loss   | -13.7    |
|    learning_rate   | 0.0001   |
|    n_updates       | 26999    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -33.2    |
| time/              |          |
|    episodes        | 36       |
|    fps             | 21       |
|    time_elapsed    | 1678     |
|    total_timesteps | 36000    |
| train/             |          |
|    actor_loss      | -31.4    |
|    critic_loss     | 0.0397   |
|    ent_coef        | 0.0451   |
|    ent_coef_loss   | -15.7    |
|    learning_rate   | 0.0001   |
|    n_updates       | 30999    |
----------------------------------
Eval num_timesteps=40000, episode_reward=-30.38 +/- 5.07
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -30.4    |
| time/              |          |
|    total_timesteps | 40000    |
| train/             |          |
|    actor_loss      | -26.8    |
|    critic_loss     | 0.0473   |
|    ent_coef        | 0.0302   |
|    ent_coef_loss   | -17.7    |
|    learning_rate   | 0.0001   |
|    n_updates       | 34999    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -32.7    |
| time/              |          |
|    episodes        | 40       |
|    fps             | 20       |
|    time_elapsed    | 1938     |
|    total_timesteps | 40000    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -32.7    |
| time/              |          |
|    episodes        | 44       |
```

```
|    episodes         | 44       |
|    fps              | 20       |
|    time_elapsed     | 2139     |
|    total_timesteps | 44000    |
| train/              |          |
|    actor_loss       | -22.5    |
|    critic_loss      | 0.0511   |
|    ent_coef         | 0.0203   |
|    ent_coef_loss    | -19.6    |
|    learning_rate    | 0.0001   |
|    n_updates        | 38999    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -32.6    |
| time/               |          |
|    episodes         | 48       |
|    fps              | 20       |
|    time_elapsed     | 2341     |
|    total_timesteps | 48000    |
| train/              |          |
|    actor_loss       | -18.6    |
|    critic_loss      | 0.032    |
|    ent_coef         | 0.0136   |
|    ent_coef_loss    | -21.6    |
|    learning_rate    | 0.0001   |
|    n_updates        | 42999    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -32.8    |
| time/               |          |
|    episodes         | 52       |
|    fps              | 20       |
|    time_elapsed     | 2547     |
|    total_timesteps | 52000    |
| train/              |          |
|    actor_loss       | -15.3    |
|    critic_loss      | 0.029    |
|    ent_coef         | 0.0091   |
|    ent_coef_loss    | -23.5    |
|    learning_rate    | 0.0001   |
|    n_updates        | 46999    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -32.7    |
| time/               |          |
|    episodes         | 56       |
|    fps              | 20       |
|    time_elapsed     | 2750     |
|    total_timesteps | 56000    |
| train/              |          |
|    actor_loss       | -12.3    |
|    critic_loss      | 0.0281   |
|    ent_coef         | 0.0061   |
|    ent_coef_loss    | -25.5    |
|    learning_rate    | 0.0001   |
|    n_updates        | 50999    |
----------------------------------
Eval num_timesteps=60000, episode_reward=-23.61 +/- 14.95
Episode length: 1000.00 +/- 0.00
```

```
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | -23.6    |
| time/               |          |
|    total_timesteps  | 60000    |
| train/              |          |
|    actor_loss       | -9.97    |
|    critic_loss      | 0.0286   |
|    ent_coef         | 0.00409  |
|    ent_coef_loss    | -27.7    |
|    learning_rate    | 0.0001   |
|    n_updates        | 54999    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -32      |
| time/               |          |
|    episodes         | 60       |
|    fps              | 19       |
|    time_elapsed     | 3012     |
|    total_timesteps  | 60000    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -31.6    |
| time/               |          |
|    episodes         | 64       |
|    fps              | 19       |
|    time_elapsed     | 3213     |
|    total_timesteps  | 64000    |
| train/              |          |
|    actor_loss       | -7.94    |
|    critic_loss      | 0.0579   |
|    ent_coef         | 0.00275  |
|    ent_coef_loss    | -28.4    |
|    learning_rate    | 0.0001   |
|    n_updates        | 58999    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -31.4    |
| time/               |          |
|    episodes         | 68       |
|    fps              | 19       |
|    time_elapsed     | 3414     |
|    total_timesteps  | 68000    |
| train/              |          |
|    actor_loss       | -6.15    |
|    critic_loss      | 0.0237   |
|    ent_coef         | 0.00188  |
|    ent_coef_loss    | -19.7    |
|    learning_rate    | 0.0001   |
|    n_updates        | 62999    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -33.9    |
| time/               |          |
|    episodes         | 72       |
```

```
|    fps              | 19       |
|    time_elapsed     | 3616     |
|    total_timesteps  | 72000    |
| train/              |          |
|    actor_loss       | -5.09    |
|    critic_loss      | 0.0184   |
|    ent_coef         | 0.00138  |
|    ent_coef_loss    | -5.69    |
|    learning_rate    | 0.0001   |
|    n_updates        | 66999    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -36.6    |
| time/               |          |
|    episodes         | 76       |
|    fps              | 19       |
|    time_elapsed     | 3822     |
|    total_timesteps  | 76000    |
| train/              |          |
|    actor_loss       | -4.06    |
|    critic_loss      | 0.0177   |
|    ent_coef         | 0.00109  |
|    ent_coef_loss    | -5.21    |
|    learning_rate    | 0.0001   |
|    n_updates        | 70999    |
----------------------------------
Eval num_timesteps=80000, episode_reward=-79.25 +/- 0.77
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | -79.3    |
| time/               |          |
|    total_timesteps  | 80000    |
| train/              |          |
|    actor_loss       | -3.54    |
|    critic_loss      | 0.0215   |
|    ent_coef         | 0.000899 |
|    ent_coef_loss    | -1.23    |
|    learning_rate    | 0.0001   |
|    n_updates        | 74999    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -38.6    |
| time/               |          |
|    episodes         | 80       |
|    fps              | 19       |
|    time_elapsed     | 4088     |
|    total_timesteps  | 80000    |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -40.5    |
| time/               |          |
|    episodes         | 84       |
|    fps              | 19       |
|    time_elapsed     | 4293     |
|    total_timesteps  | 84000    |
| train/              |          |
```

```
|    actor_loss      | -2.75    |
|    critic_loss     | 0.0146   |
|    ent_coef        | 0.000594 |
|    ent_coef_loss   | -21.9    |
|    learning_rate   | 0.0001   |
|    n_updates       | 78999    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -41.4    |
| time/              |          |
|    episodes        | 88       |
|    fps             | 19       |
|    time_elapsed    | 4496     |
|    total_timesteps | 88000    |
| train/             |          |
|    actor_loss      | -2.23    |
|    critic_loss     | 0.0152   |
|    ent_coef        | 0.000393 |
|    ent_coef_loss   | -17.8    |
|    learning_rate   | 0.0001   |
|    n_updates       | 82999    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -40.9    |
| time/              |          |
|    episodes        | 92       |
|    fps             | 19       |
|    time_elapsed    | 4700     |
|    total_timesteps | 92000    |
| train/             |          |
|    actor_loss      | -1.87    |
|    critic_loss     | 0.0114   |
|    ent_coef        | 0.000269 |
|    ent_coef_loss   | -12.5    |
|    learning_rate   | 0.0001   |
|    n_updates       | 86999    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -40.2    |
| time/              |          |
|    episodes        | 96       |
|    fps             | 19       |
|    time_elapsed    | 4904     |
|    total_timesteps | 96000    |
| train/             |          |
|    actor_loss      | -1.47    |
|    critic_loss     | 0.0153   |
|    ent_coef        | 0.000187 |
|    ent_coef_loss   | -8.67    |
|    learning_rate   | 0.0001   |
|    n_updates       | 90999    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
|    ep_rew_mean     | -42.4    |
| time/              |          |
|    episodes        | 100      |
```

```
|    fps             | 19       |
|    time_elapsed    | 5092     |
|    total_timesteps | 99713    |
| train/             |          |
|    actor_loss      | -1.05    |
|    critic_loss     | 0.00879  |
|    ent_coef        | 0.000154 |
|    ent_coef_loss   | -4.21    |
|    learning_rate   | 0.0001   |
|    n_updates       | 94712    |
----------------------------------
Eval num_timesteps=100000, episode_reward=-78.17 +/- 3.04
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -78.2    |
| time/              |          |
|    total_timesteps | 100000   |
| train/             |          |
|    actor_loss      | -1.01    |
|    critic_loss     | 0.0106   |
|    ent_coef        | 0.000154 |
|    ent_coef_loss   | 4.87     |
|    learning_rate   | 0.0001   |
|    n_updates       | 94999    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
|    ep_rew_mean     | -42.9    |
| time/              |          |
|    episodes        | 104      |
|    fps             | 19       |
|    time_elapsed    | 5358     |
|    total_timesteps | 103713   |
| train/             |          |
|    actor_loss      | -0.848   |
|    critic_loss     | 0.00764  |
|    ent_coef        | 0.000205 |
|    ent_coef_loss   | 10.5     |
|    learning_rate   | 0.0001   |
|    n_updates       | 98712    |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
|    ep_rew_mean     | -41.2    |
| time/              |          |
|    episodes        | 108      |
|    fps             | 19       |
|    time_elapsed    | 5562     |
|    total_timesteps | 107713   |
| train/             |          |
|    actor_loss      | -0.571   |
|    critic_loss     | 0.00884  |
|    ent_coef        | 0.000307 |
|    ent_coef_loss   | 9        |
|    learning_rate   | 0.0001   |
|    n_updates       | 102712   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
```

```
|    ep_rew_mean     | -36.9    |
| time/              |          |
|    episodes        | 112      |
|    fps             | 19       |
|    time_elapsed    | 5769     |
|    total_timesteps | 111713   |
| train/             |          |
|    actor_loss      | -0.523   |
|    critic_loss     | 0.00763  |
|    ent_coef        | 0.000425 |
|    ent_coef_loss   | 3.87     |
|    learning_rate   | 0.0001   |
|    n_updates       | 106712   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
|    ep_rew_mean     | -23.9    |
| time/              |          |
|    episodes        | 116      |
|    fps             | 19       |
|    time_elapsed    | 5975     |
|    total_timesteps | 115713   |
| train/             |          |
|    actor_loss      | -0.817   |
|    critic_loss     | 0.00343  |
|    ent_coef        | 0.00059  |
|    ent_coef_loss   | 5.63     |
|    learning_rate   | 0.0001   |
|    n_updates       | 110712   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
|    ep_rew_mean     | -7.96    |
| time/              |          |
|    episodes        | 120      |
|    fps             | 19       |
|    time_elapsed    | 6181     |
|    total_timesteps | 119713   |
| train/             |          |
|    actor_loss      | -0.714   |
|    critic_loss     | 0.00354  |
|    ent_coef        | 0.000821 |
|    ent_coef_loss   | 0.815    |
|    learning_rate   | 0.0001   |
|    n_updates       | 114712   |
----------------------------------
Eval num_timesteps=120000, episode_reward=524.31 +/- 302.41
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | 524      |
| time/              |          |
|    total_timesteps | 120000   |
| train/             |          |
|    actor_loss      | -0.885   |
|    critic_loss     | 0.00501  |
|    ent_coef        | 0.000836 |
|    ent_coef_loss   | 2.47     |
|    learning_rate   | 0.0001   |
|    n_updates       | 114999   |
----------------------------------
New best mean reward!
```

```
New best mean reward!
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 995      |
|    ep_rew_mean      | 18.4     |
| time/               |          |
|    episodes         | 124      |
|    fps              | 19       |
|    time_elapsed     | 6436     |
|    total_timesteps  | 123460   |
| train/              |          |
|    actor_loss       | -0.997   |
|    critic_loss      | 0.00685  |
|    ent_coef         | 0.00103  |
|    ent_coef_loss    | -3.11    |
|    learning_rate    | 0.0001   |
|    n_updates        | 118459   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 995      |
|    ep_rew_mean      | 39.5     |
| time/               |          |
|    episodes         | 128      |
|    fps              | 19       |
|    time_elapsed     | 6641     |
|    total_timesteps  | 127460   |
| train/              |          |
|    actor_loss       | -1.06    |
|    critic_loss      | 0.00446  |
|    ent_coef         | 0.00106  |
|    ent_coef_loss    | 0.716    |
|    learning_rate    | 0.0001   |
|    n_updates        | 122459   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 995      |
|    ep_rew_mean      | 66       |
| time/               |          |
|    episodes         | 132      |
|    fps              | 19       |
|    time_elapsed     | 6847     |
|    total_timesteps  | 131460   |
| train/              |          |
|    actor_loss       | -1.11    |
|    critic_loss      | 0.00618  |
|    ent_coef         | 0.00101  |
|    ent_coef_loss    | -1.65    |
|    learning_rate    | 0.0001   |
|    n_updates        | 126459   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 995      |
|    ep_rew_mean      | 86.6     |
| time/               |          |
|    episodes         | 136      |
|    fps              | 19       |
|    time_elapsed     | 7051     |
|    total_timesteps  | 135460   |
| train/              |          |
|    actor_loss       | -1.11    |
|    critic_loss      | 0.0035   |
|    ent_coef         | 0.00103  |
```

```
|    ent_coef         | 0.00103  |
|    ent_coef_loss    | 1.82     |
|    learning_rate    | 0.0001   |
|    n_updates        | 130459   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 111      |
| time/               |          |
|    episodes         | 140      |
|    fps              | 19       |
|    time_elapsed     | 7239     |
|    total_timesteps  | 139135   |
| train/              |          |
|    actor_loss       | -1.44    |
|    critic_loss      | 0.00666  |
|    ent_coef         | 0.00108  |
|    ent_coef_loss    | -2.39    |
|    learning_rate    | 0.0001   |
|    n_updates        | 134134   |
----------------------------------
Eval num_timesteps=140000, episode_reward=723.94 +/- 143.86
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | 724      |
| time/               |          |
|    total_timesteps  | 140000   |
| train/              |          |
|    actor_loss       | -1.38    |
|    critic_loss      | 0.00522  |
|    ent_coef         | 0.00105  |
|    ent_coef_loss    | -0.594   |
|    learning_rate    | 0.0001   |
|    n_updates        | 134999   |
----------------------------------
New best mean reward!
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 133      |
| time/               |          |
|    episodes         | 144      |
|    fps              | 19       |
|    time_elapsed     | 7511     |
|    total_timesteps  | 143135   |
| train/              |          |
|    actor_loss       | -1.36    |
|    critic_loss      | 0.00461  |
|    ent_coef         | 0.00118  |
|    ent_coef_loss    | 4.96     |
|    learning_rate    | 0.0001   |
|    n_updates        | 138134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 157      |
| time/               |          |
|    episodes         | 148      |
|    fps              | 19       |
|    time_elapsed     | 7715     |
|    total_timesteps  | 147135   |
```

```
| train/              |          |
|    actor_loss       | -1.39    |
|    critic_loss      | 0.00464  |
|    ent_coef         | 0.00118  |
|    ent_coef_loss    | -1.83    |
|    learning_rate    | 0.0001   |
|    n_updates        | 142134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 155      |
| time/               |          |
|    episodes         | 152      |
|    fps              | 19       |
|    time_elapsed     | 7916     |
|    total_timesteps | 151135   |
| train/              |          |
|    actor_loss       | -1.56    |
|    critic_loss      | 0.00558  |
|    ent_coef         | 0.00118  |
|    ent_coef_loss    | 1.44     |
|    learning_rate    | 0.0001   |
|    n_updates        | 146134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 158      |
| time/               |          |
|    episodes         | 156      |
|    fps              | 19       |
|    time_elapsed     | 8119     |
|    total_timesteps | 155135   |
| train/              |          |
|    actor_loss       | -1.42    |
|    critic_loss      | 0.00452  |
|    ent_coef         | 0.00121  |
|    ent_coef_loss    | -0.137   |
|    learning_rate    | 0.0001   |
|    n_updates        | 150134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 164      |
| time/               |          |
|    episodes         | 160      |
|    fps              | 19       |
|    time_elapsed     | 8321     |
|    total_timesteps | 159135   |
| train/              |          |
|    actor_loss       | -1.51    |
|    critic_loss      | 0.00802  |
|    ent_coef         | 0.00138  |
|    ent_coef_loss    | 0.493    |
|    learning_rate    | 0.0001   |
|    n_updates        | 154134   |
----------------------------------
Eval num_timesteps=160000, episode_reward=296.36 +/- 331.31
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
```

```
|    mean_reward     | 296      |
| time/              |          |
|    total_timesteps | 160000   |
| train/             |          |
|    actor_loss      | -1.66    |
|    critic_loss     | 0.00581  |
|    ent_coef        | 0.00139  |
|    ent_coef_loss   | -1.45    |
|    learning_rate   | 0.0001   |
|    n_updates       | 154999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 991      |
|    ep_rew_mean     | 170      |
| time/              |          |
|    episodes        | 164      |
|    fps             | 18       |
|    time_elapsed    | 8586     |
|    total_timesteps | 163135   |
| train/             |          |
|    actor_loss      | -1.71    |
|    critic_loss     | 0.00602  |
|    ent_coef        | 0.00137  |
|    ent_coef_loss   | -0.711   |
|    learning_rate   | 0.0001   |
|    n_updates       | 158134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 991      |
|    ep_rew_mean     | 179      |
| time/              |          |
|    episodes        | 168      |
|    fps             | 19       |
|    time_elapsed    | 8788     |
|    total_timesteps | 167135   |
| train/             |          |
|    actor_loss      | -1.68    |
|    critic_loss     | 0.00392  |
|    ent_coef        | 0.00126  |
|    ent_coef_loss   | -2.15    |
|    learning_rate   | 0.0001   |
|    n_updates       | 162134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 991      |
|    ep_rew_mean     | 185      |
| time/              |          |
|    episodes        | 172      |
|    fps             | 19       |
|    time_elapsed    | 8990     |
|    total_timesteps | 171135   |
| train/             |          |
|    actor_loss      | -1.67    |
|    critic_loss     | 0.00357  |
|    ent_coef        | 0.00112  |
|    ent_coef_loss   | 0.252    |
|    learning_rate   | 0.0001   |
|    n_updates       | 166134   |
----------------------------------
----------------------------------
| rollout/           |          |
```

```
|    ep_len_mean     | 991      |
|    ep_rew_mean     | 190      |
| time/              |          |
|    episodes        | 176      |
|    fps             | 19       |
|    time_elapsed    | 9195     |
|    total_timesteps | 175135   |
| train/             |          |
|    actor_loss      | -1.64    |
|    critic_loss     | 0.0082   |
|    ent_coef        | 0.000987 |
|    ent_coef_loss   | -2.36    |
|    learning_rate   | 0.0001   |
|    n_updates       | 170134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 991      |
|    ep_rew_mean     | 190      |
| time/              |          |
|    episodes        | 180      |
|    fps             | 19       |
|    time_elapsed    | 9399     |
|    total_timesteps | 179135   |
| train/             |          |
|    actor_loss      | -1.57    |
|    critic_loss     | 0.00375  |
|    ent_coef        | 0.000918 |
|    ent_coef_loss   | 1.24     |
|    learning_rate   | 0.0001   |
|    n_updates       | 174134   |
----------------------------------
Eval num_timesteps=180000, episode_reward=-85.45 +/- 6.25
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -85.5    |
| time/              |          |
|    total_timesteps | 180000   |
| train/             |          |
|    actor_loss      | -1.57    |
|    critic_loss     | 0.00359  |
|    ent_coef        | 0.000938 |
|    ent_coef_loss   | 2.82     |
|    learning_rate   | 0.0001   |
|    n_updates       | 174999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 991      |
|    ep_rew_mean     | 190      |
| time/              |          |
|    episodes        | 184      |
|    fps             | 18       |
|    time_elapsed    | 9661     |
|    total_timesteps | 183135   |
| train/             |          |
|    actor_loss      | -1.44    |
|    critic_loss     | 0.00354  |
|    ent_coef        | 0.00103  |
|    ent_coef_loss   | -1.88    |
|    learning_rate   | 0.0001   |
|    n_updates       | 178134   |
```

```
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 207      |
| time/               |          |
|    episodes         | 188      |
|    fps              | 18       |
|    time_elapsed     | 9865     |
|    total_timesteps  | 187135   |
| train/              |          |
|    actor_loss       | -1.48    |
|    critic_loss      | 0.00452  |
|    ent_coef         | 0.000923 |
|    ent_coef_loss    | -0.992   |
|    learning_rate    | 0.0001   |
|    n_updates        | 182134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 233      |
| time/               |          |
|    episodes         | 192      |
|    fps              | 18       |
|    time_elapsed     | 10069    |
|    total_timesteps  | 191135   |
| train/              |          |
|    actor_loss       | -1.57    |
|    critic_loss      | 0.00411  |
|    ent_coef         | 0.000886 |
|    ent_coef_loss    | 1.74     |
|    learning_rate    | 0.0001   |
|    n_updates        | 186134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 991      |
|    ep_rew_mean      | 261      |
| time/               |          |
|    episodes         | 196      |
|    fps              | 18       |
|    time_elapsed     | 10273    |
|    total_timesteps  | 195135   |
| train/              |          |
|    actor_loss       | -1.53    |
|    critic_loss      | 0.0056   |
|    ent_coef         | 0.000933 |
|    ent_coef_loss    | 2.89     |
|    learning_rate    | 0.0001   |
|    n_updates        | 190134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 994      |
|    ep_rew_mean      | 285      |
| time/               |          |
|    episodes         | 200      |
|    fps              | 19       |
|    time_elapsed     | 10478    |
|    total_timesteps  | 199135   |
| train/              |          |
|    actor_loss       | -1.49    |
|    critic_loss      | 0.00406  |
|    ent_coef         | 0.000988 |
```

```
|    ent_coef         | 0.000988 |
|    ent_coef_loss    | 2.23     |
|    learning_rate    | 0.0001   |
|    n_updates        | 194134   |
----------------------------------
Eval num_timesteps=200000, episode_reward=-11.20 +/- 151.92
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | -11.2    |
| time/               |          |
|    total_timesteps  | 200000   |
| train/              |          |
|    actor_loss       | -1.52    |
|    critic_loss      | 0.00888  |
|    ent_coef         | 0.000974 |
|    ent_coef_loss    | 2.47     |
|    learning_rate    | 0.0001   |
|    n_updates        | 194999   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 994      |
|    ep_rew_mean      | 292      |
| time/               |          |
|    episodes         | 204      |
|    fps              | 18       |
|    time_elapsed     | 10745    |
|    total_timesteps  | 203135   |
| train/              |          |
|    actor_loss       | -1.62    |
|    critic_loss      | 0.00438  |
|    ent_coef         | 0.000963 |
|    ent_coef_loss    | -1.91    |
|    learning_rate    | 0.0001   |
|    n_updates        | 198134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 994      |
|    ep_rew_mean      | 293      |
| time/               |          |
|    episodes         | 208      |
|    fps              | 18       |
|    time_elapsed     | 10949    |
|    total_timesteps  | 207135   |
| train/              |          |
|    actor_loss       | -1.45    |
|    critic_loss      | 0.00763  |
|    ent_coef         | 0.00107  |
|    ent_coef_loss    | -4.41    |
|    learning_rate    | 0.0001   |
|    n_updates        | 202134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 994      |
|    ep_rew_mean      | 286      |
| time/               |          |
|    episodes         | 212      |
|    fps              | 18       |
|    time_elapsed     | 11152    |
|    total_timesteps  | 211135   |
| train/              |          |
```

```
| train/              |          |
|    actor_loss       | -1.99    |
|    critic_loss      | 0.0185   |
|    ent_coef         | 0.00115  |
|    ent_coef_loss    | 2.75     |
|    learning_rate    | 0.0001   |
|    n_updates        | 206134   |
---------------------------------
---------------------------------
| rollout/            |          |
|    ep_len_mean      | 994      |
|    ep_rew_mean      | 280      |
| time/               |          |
|    episodes         | 216      |
|    fps              | 18       |
|    time_elapsed     | 11355    |
|    total_timesteps  | 215135   |
| train/              |          |
|    actor_loss       | -1.87    |
|    critic_loss      | 0.0143   |
|    ent_coef         | 0.00123  |
|    ent_coef_loss    | 2.39     |
|    learning_rate    | 0.0001   |
|    n_updates        | 210134   |
---------------------------------
---------------------------------
| rollout/            |          |
|    ep_len_mean      | 994      |
|    ep_rew_mean      | 262      |
| time/               |          |
|    episodes         | 220      |
|    fps              | 18       |
|    time_elapsed     | 11558    |
|    total_timesteps  | 219135   |
| train/              |          |
|    actor_loss       | -2.16    |
|    critic_loss      | 0.00917  |
|    ent_coef         | 0.00128  |
|    ent_coef_loss    | 0.064    |
|    learning_rate    | 0.0001   |
|    n_updates        | 214134   |
---------------------------------
Eval num_timesteps=220000, episode_reward=-88.99 +/- 2.04
Episode length: 1000.00 +/- 0.00
---------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | -89      |
| time/               |          |
|    total_timesteps  | 220000   |
| train/              |          |
|    actor_loss       | -2.02    |
|    critic_loss      | 0.0111   |
|    ent_coef         | 0.00127  |
|    ent_coef_loss    | -1.2     |
|    learning_rate    | 0.0001   |
|    n_updates        | 214999   |
---------------------------------
---------------------------------
| rollout/            |          |
|    ep_len_mean      | 997      |
|    ep_rew_mean      | 233      |
| time/               |          |
|    episodes         | 224      |
|    fps              | 18       |
```

```
|    time_elapsed    | 11822    |
|    total_timesteps | 223135   |
| train/             |          |
|    actor_loss      | -2.23    |
|    critic_loss     | 0.0242   |
|    ent_coef        | 0.00117  |
|    ent_coef_loss   | 0.81     |
|    learning_rate   | 0.0001   |
|    n_updates       | 218134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
|    ep_rew_mean     | 210      |
| time/              |          |
|    episodes        | 228      |
|    fps             | 18       |
|    time_elapsed    | 12022    |
|    total_timesteps | 227135   |
| train/             |          |
|    actor_loss      | -1.96    |
|    critic_loss     | 0.0152   |
|    ent_coef        | 0.00112  |
|    ent_coef_loss   | -0.293   |
|    learning_rate   | 0.0001   |
|    n_updates       | 222134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
|    ep_rew_mean     | 182      |
| time/              |          |
|    episodes        | 232      |
|    fps             | 18       |
|    time_elapsed    | 12226    |
|    total_timesteps | 231135   |
| train/             |          |
|    actor_loss      | -2.4     |
|    critic_loss     | 0.033    |
|    ent_coef        | 0.00114  |
|    ent_coef_loss   | 0.47     |
|    learning_rate   | 0.0001   |
|    n_updates       | 226134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 997      |
|    ep_rew_mean     | 159      |
| time/              |          |
|    episodes        | 236      |
|    fps             | 18       |
|    time_elapsed    | 12428    |
|    total_timesteps | 235135   |
| train/             |          |
|    actor_loss      | -2.49    |
|    critic_loss     | 0.0396   |
|    ent_coef        | 0.00121  |
|    ent_coef_loss   | -0.673   |
|    learning_rate   | 0.0001   |
|    n_updates       | 230134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
```

```
|    ep_rew_mean     | 134      |
| time/              |          |
|    episodes        | 240      |
|    fps             | 18       |
|    time_elapsed    | 12631    |
|    total_timesteps | 239135   |
| train/             |          |
|    actor_loss      | -2.75    |
|    critic_loss     | 0.0658   |
|    ent_coef        | 0.00137  |
|    ent_coef_loss   | -1.23    |
|    learning_rate   | 0.0001   |
|    n_updates       | 234134   |
----------------------------------
Eval num_timesteps=240000, episode_reward=-77.94 +/- 17.44
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -77.9    |
| time/              |          |
|    total_timesteps | 240000   |
| train/             |          |
|    actor_loss      | -2.2     |
|    critic_loss     | 0.033    |
|    ent_coef        | 0.00139  |
|    ent_coef_loss   | -0.617   |
|    learning_rate   | 0.0001   |
|    n_updates       | 234999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 111      |
| time/              |          |
|    episodes        | 244      |
|    fps             | 18       |
|    time_elapsed    | 12895    |
|    total_timesteps | 243135   |
| train/             |          |
|    actor_loss      | -2.94    |
|    critic_loss     | 0.0379   |
|    ent_coef        | 0.00145  |
|    ent_coef_loss   | 0.616    |
|    learning_rate   | 0.0001   |
|    n_updates       | 238134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 84.7     |
| time/              |          |
|    episodes        | 248      |
|    fps             | 18       |
|    time_elapsed    | 13098    |
|    total_timesteps | 247135   |
| train/             |          |
|    actor_loss      | -2.41    |
|    critic_loss     | 0.0642   |
|    ent_coef        | 0.00167  |
|    ent_coef_loss   | -3.13    |
|    learning_rate   | 0.0001   |
|    n_updates       | 242134   |
----------------------------------
```

```
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | 85.3     |
| time/               |          |
|    episodes         | 252      |
|    fps              | 18       |
|    time_elapsed     | 13301    |
|    total_timesteps  | 251135   |
| train/              |          |
|    actor_loss       | -2.93    |
|    critic_loss      | 0.0948   |
|    ent_coef         | 0.00206  |
|    ent_coef_loss    | 2.89     |
|    learning_rate    | 0.0001   |
|    n_updates        | 246134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | 80.9     |
| time/               |          |
|    episodes         | 256      |
|    fps              | 18       |
|    time_elapsed     | 13504    |
|    total_timesteps  | 255135   |
| train/              |          |
|    actor_loss       | -3.02    |
|    critic_loss      | 0.0886   |
|    ent_coef         | 0.0026   |
|    ent_coef_loss    | -0.217   |
|    learning_rate    | 0.0001   |
|    n_updates        | 250134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | 74.7     |
| time/               |          |
|    episodes         | 260      |
|    fps              | 18       |
|    time_elapsed     | 13707    |
|    total_timesteps  | 259135   |
| train/              |          |
|    actor_loss       | -3.55    |
|    critic_loss      | 0.115    |
|    ent_coef         | 0.00289  |
|    ent_coef_loss    | 0.117    |
|    learning_rate    | 0.0001   |
|    n_updates        | 254134   |
----------------------------------
Eval num_timesteps=260000, episode_reward=-53.58 +/- 2.52
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | -53.6    |
| time/               |          |
|    total_timesteps  | 260000   |
| train/              |          |
|    actor_loss       | -4.2     |
|    critic_loss      | 0.493    |
|    ent_coef         | 0.00302  |
|    ent_coef_loss    | 1.16     |
|    learning_rate    | 0.0001   |
```

```
|    learning_rate   | 0.0001   |
|    n_updates       | 254999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 66.9     |
| time/              |          |
|    episodes        | 264      |
|    fps             | 18       |
|    time_elapsed    | 13973    |
|    total_timesteps | 263135   |
| train/             |          |
|    actor_loss      | -4.05    |
|    critic_loss     | 0.0581   |
|    ent_coef        | 0.00348  |
|    ent_coef_loss   | 0.947    |
|    learning_rate   | 0.0001   |
|    n_updates       | 258134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 57.3     |
| time/              |          |
|    episodes        | 268      |
|    fps             | 18       |
|    time_elapsed    | 14177    |
|    total_timesteps | 267135   |
| train/             |          |
|    actor_loss      | -4.72    |
|    critic_loss     | 0.0577   |
|    ent_coef        | 0.00345  |
|    ent_coef_loss   | -1.24    |
|    learning_rate   | 0.0001   |
|    n_updates       | 262134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 51.3     |
| time/              |          |
|    episodes        | 272      |
|    fps             | 18       |
|    time_elapsed    | 14378    |
|    total_timesteps | 271135   |
| train/             |          |
|    actor_loss      | -4.42    |
|    critic_loss     | 0.0845   |
|    ent_coef        | 0.0033   |
|    ent_coef_loss   | 1        |
|    learning_rate   | 0.0001   |
|    n_updates       | 266134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 46.8     |
| time/              |          |
|    episodes        | 276      |
|    fps             | 18       |
|    time_elapsed    | 14582    |
|    total_timesteps | 275135   |
| train/             |          |
|    actor_loss      | -4.71    |
```

```
|    actor_loss      | -4.71    |
|    critic_loss     | 0.341    |
|    ent_coef        | 0.00302  |
|    ent_coef_loss   | 0.352    |
|    learning_rate   | 0.0001   |
|    n_updates       | 270134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 47.6     |
| time/              |          |
|    episodes        | 280      |
|    fps             | 18       |
|    time_elapsed    | 14785    |
|    total_timesteps | 279135   |
| train/             |          |
|    actor_loss      | -5.81    |
|    critic_loss     | 0.105    |
|    ent_coef        | 0.00309  |
|    ent_coef_loss   | 0.34     |
|    learning_rate   | 0.0001   |
|    n_updates       | 274134   |
----------------------------------
Eval num_timesteps=280000, episode_reward=-60.06 +/- 10.65
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -60.1    |
| time/              |          |
|    total_timesteps | 280000   |
| train/             |          |
|    actor_loss      | -5.43    |
|    critic_loss     | 0.63     |
|    ent_coef        | 0.00303  |
|    ent_coef_loss   | -0.336   |
|    learning_rate   | 0.0001   |
|    n_updates       | 274999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 48       |
| time/              |          |
|    episodes        | 284      |
|    fps             | 18       |
|    time_elapsed    | 15051    |
|    total_timesteps | 283135   |
| train/             |          |
|    actor_loss      | -4.95    |
|    critic_loss     | 0.15     |
|    ent_coef        | 0.00321  |
|    ent_coef_loss   | -0.114   |
|    learning_rate   | 0.0001   |
|    n_updates       | 278134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 29.9     |
| time/              |          |
|    episodes        | 288      |
|    fps             | 18       |
|    time_elapsed    | 15254    |
```

```
|    time_elapsed    | 15254    |
|    total_timesteps | 287135   |
| train/             |          |
|    actor_loss      | -5.79    |
|    critic_loss     | 0.302    |
|    ent_coef        | 0.00326  |
|    ent_coef_loss   | -1.66    |
|    learning_rate   | 0.0001   |
|    n_updates       | 282134   |
---------------------------------
---------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | 2.95     |
| time/              |          |
|    episodes        | 292      |
|    fps             | 18       |
|    time_elapsed    | 15457    |
|    total_timesteps | 291135   |
| train/             |          |
|    actor_loss      | -5.39    |
|    critic_loss     | 0.156    |
|    ent_coef        | 0.00324  |
|    ent_coef_loss   | 0.902    |
|    learning_rate   | 0.0001   |
|    n_updates       | 286134   |
---------------------------------
---------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -25.9    |
| time/              |          |
|    episodes        | 296      |
|    fps             | 18       |
|    time_elapsed    | 15658    |
|    total_timesteps | 295135   |
| train/             |          |
|    actor_loss      | -5.48    |
|    critic_loss     | 0.308    |
|    ent_coef        | 0.00342  |
|    ent_coef_loss   | 1.86     |
|    learning_rate   | 0.0001   |
|    n_updates       | 290134   |
---------------------------------
---------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -49.2    |
| time/              |          |
|    episodes        | 300      |
|    fps             | 18       |
|    time_elapsed    | 15861    |
|    total_timesteps | 299135   |
| train/             |          |
|    actor_loss      | -6.06    |
|    critic_loss     | 0.115    |
|    ent_coef        | 0.00372  |
|    ent_coef_loss   | -0.585   |
|    learning_rate   | 0.0001   |
|    n_updates       | 294134   |
---------------------------------
Eval num_timesteps=300000, episode_reward=-73.17 +/- 11.71
Episode length: 1000.00 +/- 0.00
---------------------------------
| eval/              |          |
```

```
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -73.2    |
| time/              |          |
|    total_timesteps | 300000   |
| train/             |          |
|    actor_loss      | -5.34    |
|    critic_loss     | 0.34     |
|    ent_coef        | 0.00378  |
|    ent_coef_loss   | -0.41    |
|    learning_rate   | 0.0001   |
|    n_updates       | 294999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -57.9    |
| time/              |          |
|    episodes        | 304      |
|    fps             | 18       |
|    time_elapsed    | 16131    |
|    total_timesteps | 303135   |
| train/             |          |
|    actor_loss      | -6.56    |
|    critic_loss     | 0.23     |
|    ent_coef        | 0.00387  |
|    ent_coef_loss   | 1.09     |
|    learning_rate   | 0.0001   |
|    n_updates       | 298134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -60.5    |
| time/              |          |
|    episodes        | 308      |
|    fps             | 18       |
|    time_elapsed    | 16334    |
|    total_timesteps | 307135   |
| train/             |          |
|    actor_loss      | -5.9     |
|    critic_loss     | 0.279    |
|    ent_coef        | 0.00413  |
|    ent_coef_loss   | 0.17     |
|    learning_rate   | 0.0001   |
|    n_updates       | 302134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -59.6    |
| time/              |          |
|    episodes        | 312      |
|    fps             | 18       |
|    time_elapsed    | 16537    |
|    total_timesteps | 311135   |
| train/             |          |
|    actor_loss      | -5.95    |
|    critic_loss     | 0.384    |
|    ent_coef        | 0.00457  |
|    ent_coef_loss   | -0.825   |
|    learning_rate   | 0.0001   |
|    n_updates       | 306134   |
----------------------------------
----------------------------------
```

```
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -67.8    |
| time/              |          |
|    episodes        | 316      |
|    fps             | 18       |
|    time_elapsed    | 16740    |
|    total_timesteps | 315135   |
| train/             |          |
|    actor_loss      | -6.62    |
|    critic_loss     | 0.309    |
|    ent_coef        | 0.00524  |
|    ent_coef_loss   | 1.54     |
|    learning_rate   | 0.0001   |
|    n_updates       | 310134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -66.9    |
| time/              |          |
|    episodes        | 320      |
|    fps             | 18       |
|    time_elapsed    | 16944    |
|    total_timesteps | 319135   |
| train/             |          |
|    actor_loss      | -7.4     |
|    critic_loss     | 0.682    |
|    ent_coef        | 0.00594  |
|    ent_coef_loss   | 1.52     |
|    learning_rate   | 0.0001   |
|    n_updates       | 314134   |
----------------------------------
Eval num_timesteps=320000, episode_reward=-48.35 +/- 28.16
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -48.4    |
| time/              |          |
|    total_timesteps | 320000   |
| train/             |          |
|    actor_loss      | -6.83    |
|    critic_loss     | 0.406    |
|    ent_coef        | 0.00615  |
|    ent_coef_loss   | -0.399   |
|    learning_rate   | 0.0001   |
|    n_updates       | 314999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -64.8    |
| time/              |          |
|    episodes        | 324      |
|    fps             | 18       |
|    time_elapsed    | 17211    |
|    total_timesteps | 323135   |
| train/             |          |
|    actor_loss      | -7.87    |
|    critic_loss     | 0.589    |
|    ent_coef        | 0.00677  |
|    ent_coef_loss   | 0.725    |
|    learning_rate   | 0.0001   |
```

```
|    n_updates       | 318134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -62.8    |
| time/              |          |
|    episodes        | 328      |
|    fps             | 18       |
|    time_elapsed    | 17413    |
|    total_timesteps | 327135   |
| train/             |          |
|    actor_loss      | -7.9     |
|    critic_loss     | 0.385    |
|    ent_coef        | 0.00713  |
|    ent_coef_loss   | -0.718   |
|    learning_rate   | 0.0001   |
|    n_updates       | 322134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -60.1    |
| time/              |          |
|    episodes        | 332      |
|    fps             | 18       |
|    time_elapsed    | 17613    |
|    total_timesteps | 331135   |
| train/             |          |
|    actor_loss      | -9.73    |
|    critic_loss     | 0.387    |
|    ent_coef        | 0.00691  |
|    ent_coef_loss   | -1.62    |
|    learning_rate   | 0.0001   |
|    n_updates       | 326134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -58.2    |
| time/              |          |
|    episodes        | 336      |
|    fps             | 18       |
|    time_elapsed    | 17816    |
|    total_timesteps | 335135   |
| train/             |          |
|    actor_loss      | -8.52    |
|    critic_loss     | 0.426    |
|    ent_coef        | 0.00607  |
|    ent_coef_loss   | 0.907    |
|    learning_rate   | 0.0001   |
|    n_updates       | 330134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -55.6    |
| time/              |          |
|    episodes        | 340      |
|    fps             | 18       |
|    time_elapsed    | 18018    |
|    total_timesteps | 339135   |
| train/             |          |
|    actor_loss      | -7.57    |
```

```
|    critic_loss     | 0.355    |
|    ent_coef        | 0.00637  |
|    ent_coef_loss   | -1.84    |
|    learning_rate   | 0.0001   |
|    n_updates       | 334134   |
----------------------------------
Eval num_timesteps=340000, episode_reward=-59.16 +/- 8.49
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |           |
|    mean_ep_length  | 1e+03     |
|    mean_reward     | -59.2     |
| time/              |           |
|    total_timesteps | 340000    |
| train/             |           |
|    actor_loss      | -8.04     |
|    critic_loss     | 0.463     |
|    ent_coef        | 0.00622   |
|    ent_coef_loss   | 0.527     |
|    learning_rate   | 0.0001    |
|    n_updates       | 334999    |
----------------------------------
----------------------------------
| rollout/           |           |
|    ep_len_mean     | 1e+03     |
|    ep_rew_mean     | -52       |
| time/              |           |
|    episodes        | 344       |
|    fps             | 18        |
|    time_elapsed    | 18285     |
|    total_timesteps | 343135    |
| train/             |           |
|    actor_loss      | -9.57     |
|    critic_loss     | 0.629     |
|    ent_coef        | 0.00557   |
|    ent_coef_loss   | -0.654    |
|    learning_rate   | 0.0001    |
|    n_updates       | 338134    |
----------------------------------
----------------------------------
| rollout/           |           |
|    ep_len_mean     | 1e+03     |
|    ep_rew_mean     | -48.7     |
| time/              |           |
|    episodes        | 348       |
|    fps             | 18        |
|    time_elapsed    | 18488     |
|    total_timesteps | 347135    |
| train/             |           |
|    actor_loss      | -8.45     |
|    critic_loss     | 0.542     |
|    ent_coef        | 0.00537   |
|    ent_coef_loss   | -1.55     |
|    learning_rate   | 0.0001    |
|    n_updates       | 342134    |
----------------------------------
----------------------------------
| rollout/           |           |
|    ep_len_mean     | 1e+03     |
|    ep_rew_mean     | -45.4     |
| time/              |           |
|    episodes        | 352       |
|    fps             | 18        |
|    time_elapsed    | 18690     |
|    total_timesteps | 351135    |
```

```
|    total_timesteps | 351135   |
| train/              |          |
|    actor_loss       | -11      |
|    critic_loss      | 0.866    |
|    ent_coef         | 0.00598  |
|    ent_coef_loss    | -0.71    |
|    learning_rate    | 0.0001   |
|    n_updates        | 346134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -46      |
| time/               |          |
|    episodes         | 356      |
|    fps              | 18       |
|    time_elapsed     | 18893    |
|    total_timesteps | 355135   |
| train/              |          |
|    actor_loss       | -8.39    |
|    critic_loss      | 0.492    |
|    ent_coef         | 0.0058   |
|    ent_coef_loss    | -0.649   |
|    learning_rate    | 0.0001   |
|    n_updates        | 350134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -46.5    |
| time/               |          |
|    episodes         | 360      |
|    fps              | 18       |
|    time_elapsed     | 19098    |
|    total_timesteps | 359135   |
| train/              |          |
|    actor_loss       | -9.98    |
|    critic_loss      | 0.846    |
|    ent_coef         | 0.00601  |
|    ent_coef_loss    | -0.16    |
|    learning_rate    | 0.0001   |
|    n_updates        | 354134   |
----------------------------------
Eval num_timesteps=360000, episode_reward=-56.11 +/- 15.32
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | -56.1    |
| time/               |          |
|    total_timesteps | 360000   |
| train/              |          |
|    actor_loss       | -12.1    |
|    critic_loss      | 4.39     |
|    ent_coef         | 0.00589  |
|    ent_coef_loss    | -2.49    |
|    learning_rate    | 0.0001   |
|    n_updates        | 354999   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -46.9    |
| time/               |          |
|    episodes         | 364      |
```

```
|    episodes         | 364      |
|    fps              | 18       |
|    time_elapsed     | 19363    |
|    total_timesteps  | 363135   |
| train/              |          |
|    actor_loss       | -9.24    |
|    critic_loss      | 1.48     |
|    ent_coef         | 0.006    |
|    ent_coef_loss    | -0.277   |
|    learning_rate    | 0.0001   |
|    n_updates        | 358134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -46.7    |
| time/               |          |
|    episodes         | 368      |
|    fps              | 18       |
|    time_elapsed     | 19565    |
|    total_timesteps  | 367135   |
| train/              |          |
|    actor_loss       | -9.34    |
|    critic_loss      | 2.01     |
|    ent_coef         | 0.0063   |
|    ent_coef_loss    | 0.286    |
|    learning_rate    | 0.0001   |
|    n_updates        | 362134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -45.8    |
| time/               |          |
|    episodes         | 372      |
|    fps              | 18       |
|    time_elapsed     | 19767    |
|    total_timesteps  | 371135   |
| train/              |          |
|    actor_loss       | -8.9     |
|    critic_loss      | 1.31     |
|    ent_coef         | 0.00719  |
|    ent_coef_loss    | -0.749   |
|    learning_rate    | 0.0001   |
|    n_updates        | 366134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -45.4    |
| time/               |          |
|    episodes         | 376      |
|    fps              | 18       |
|    time_elapsed     | 19970    |
|    total_timesteps  | 375135   |
| train/              |          |
|    actor_loss       | -12      |
|    critic_loss      | 14.4     |
|    ent_coef         | 0.00828  |
|    ent_coef_loss    | 2.33     |
|    learning_rate    | 0.0001   |
|    n_updates        | 370134   |
----------------------------------
----------------------------------
| rollout/            |          |
```

```
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -45.5    |
| time/              |          |
|    episodes        | 380      |
|    fps             | 18       |
|    time_elapsed    | 20172    |
|    total_timesteps | 379135   |
| train/             |          |
|    actor_loss      | -11.3    |
|    critic_loss     | 4.01     |
|    ent_coef        | 0.00893  |
|    ent_coef_loss   | 1.5      |
|    learning_rate   | 0.0001   |
|    n_updates       | 374134   |
----------------------------------
Eval num_timesteps=380000, episode_reward=-78.72 +/- 12.76
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -78.7    |
| time/              |          |
|    total_timesteps | 380000   |
| train/             |          |
|    actor_loss      | -9.7     |
|    critic_loss     | 3.74     |
|    ent_coef        | 0.00913  |
|    ent_coef_loss   | 1.16     |
|    learning_rate   | 0.0001   |
|    n_updates       | 374999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -45.2    |
| time/              |          |
|    episodes        | 384      |
|    fps             | 18       |
|    time_elapsed    | 20436    |
|    total_timesteps | 383135   |
| train/             |          |
|    actor_loss      | -13      |
|    critic_loss     | 5.46     |
|    ent_coef        | 0.0102   |
|    ent_coef_loss   | 0.19     |
|    learning_rate   | 0.0001   |
|    n_updates       | 378134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -45.4    |
| time/              |          |
|    episodes        | 388      |
|    fps             | 18       |
|    time_elapsed    | 20639    |
|    total_timesteps | 387135   |
| train/             |          |
|    actor_loss      | -21.5    |
|    critic_loss     | 15.8     |
|    ent_coef        | 0.0131   |
|    ent_coef_loss   | -0.185   |
|    learning_rate   | 0.0001   |
|    n_updates       | 382134   |
```

```
---------------------------------
---------------------------------
| rollout/            |         |
|    ep_len_mean      | 1e+03   |
|    ep_rew_mean      | -46.5   |
| time/               |         |
|    episodes         | 392     |
|    fps              | 18      |
|    time_elapsed     | 20840   |
|    total_timesteps  | 391135  |
| train/              |         |
|    actor_loss       | -22.4   |
|    critic_loss      | 15.2    |
|    ent_coef         | 0.0163  |
|    ent_coef_loss    | 0.68    |
|    learning_rate    | 0.0001  |
|    n_updates        | 386134  |
---------------------------------
---------------------------------
| rollout/            |         |
|    ep_len_mean      | 1e+03   |
|    ep_rew_mean      | -47.5   |
| time/               |         |
|    episodes         | 396     |
|    fps              | 18      |
|    time_elapsed     | 21048   |
|    total_timesteps  | 395135  |
| train/              |         |
|    actor_loss       | -36.3   |
|    critic_loss      | 38.6    |
|    ent_coef         | 0.021   |
|    ent_coef_loss    | 1.78    |
|    learning_rate    | 0.0001  |
|    n_updates        | 390134  |
---------------------------------
---------------------------------
| rollout/            |         |
|    ep_len_mean      | 1e+03   |
|    ep_rew_mean      | -47.8   |
| time/               |         |
|    episodes         | 400     |
|    fps              | 18      |
|    time_elapsed     | 21255   |
|    total_timesteps  | 399135  |
| train/              |         |
|    actor_loss       | -26     |
|    critic_loss      | 29.2    |
|    ent_coef         | 0.024   |
|    ent_coef_loss    | 1.55    |
|    learning_rate    | 0.0001  |
|    n_updates        | 394134  |
---------------------------------
Eval num_timesteps=400000, episode_reward=-81.05 +/- 4.42
Episode length: 1000.00 +/- 0.00
---------------------------------
| eval/               |         |
|    mean_ep_length   | 1e+03   |
|    mean_reward      | -81.1   |
| time/               |         |
|    total_timesteps  | 400000  |
| train/              |         |
|    actor_loss       | -18.2   |
|    critic_loss      | 174     |
|    ent_coef         | 0.0258  |
```

```
|    ent_coef_loss   | 0.987    |
|    learning_rate   | 0.0001   |
|    n_updates       | 394999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -48.1    |
| time/              |          |
|    episodes        | 404      |
|    fps             | 18       |
|    time_elapsed    | 21524    |
|    total_timesteps | 403135   |
| train/             |          |
|    actor_loss      | -44.6    |
|    critic_loss     | 133      |
|    ent_coef        | 0.0306   |
|    ent_coef_loss   | 1        |
|    learning_rate   | 0.0001   |
|    n_updates       | 398134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -49      |
| time/              |          |
|    episodes        | 408      |
|    fps             | 18       |
|    time_elapsed    | 21729    |
|    total_timesteps | 407135   |
| train/             |          |
|    actor_loss      | -49.7    |
|    critic_loss     | 168      |
|    ent_coef        | 0.0353   |
|    ent_coef_loss   | 0.691    |
|    learning_rate   | 0.0001   |
|    n_updates       | 402134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -49.4    |
| time/              |          |
|    episodes        | 412      |
|    fps             | 18       |
|    time_elapsed    | 21934    |
|    total_timesteps | 411135   |
| train/             |          |
|    actor_loss      | -60.8    |
|    critic_loss     | 129      |
|    ent_coef        | 0.0407   |
|    ent_coef_loss   | 0.00503  |
|    learning_rate   | 0.0001   |
|    n_updates       | 406134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -49.5    |
| time/              |          |
|    episodes        | 416      |
|    fps             | 18       |
|    time_elapsed    | 22138    |
|    total_timesteps | 415135   |
```

```
| train/              |          |
|    actor_loss       | -47.2    |
|    critic_loss      | 67.1     |
|    ent_coef         | 0.0402   |
|    ent_coef_loss    | -1.49    |
|    learning_rate    | 0.0001   |
|    n_updates        | 410134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -50.1    |
| time/               |          |
|    episodes         | 420      |
|    fps              | 18       |
|    time_elapsed     | 22342    |
|    total_timesteps  | 419135   |
| train/              |          |
|    actor_loss       | -56      |
|    critic_loss      | 168      |
|    ent_coef         | 0.0448   |
|    ent_coef_loss    | 0.622    |
|    learning_rate    | 0.0001   |
|    n_updates        | 414134   |
----------------------------------
Eval num_timesteps=420000, episode_reward=-73.35 +/- 8.27
----------------------------------
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | -73.4    |
| time/               |          |
|    total_timesteps  | 420000   |
| train/              |          |
|    actor_loss       | -50.5    |
|    critic_loss      | 122      |
|    ent_coef         | 0.0452   |
|    ent_coef_loss    | -0.676   |
|    learning_rate    | 0.0001   |
|    n_updates        | 414999   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -52      |
| time/               |          |
|    episodes         | 424      |
|    fps              | 18       |
|    time_elapsed     | 22613    |
|    total_timesteps  | 423135   |
| train/              |          |
|    actor_loss       | -56.2    |
|    critic_loss      | 1.52e+03 |
|    ent_coef         | 0.0484   |
|    ent_coef_loss    | -0.323   |
|    learning_rate    | 0.0001   |
|    n_updates        | 418134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -53.9    |
| time/               |          |
|    episodes         | 428      |
|    fps              | 18       |
```

```
|    time_elapsed    | 22819    |
|    total_timesteps | 427135   |
| train/             |          |
|    actor_loss      | -67.8    |
|    critic_loss     | 72       |
|    ent_coef        | 0.0465   |
|    ent_coef_loss   | -0.389   |
|    learning_rate   | 0.0001   |
|    n_updates       | 422134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -56.2    |
| time/              |          |
|    episodes        | 432      |
|    fps             | 18       |
|    time_elapsed    | 23027    |
|    total_timesteps | 431135   |
| train/             |          |
|    actor_loss      | -70.6    |
|    critic_loss     | 305      |
|    ent_coef        | 0.0517   |
|    ent_coef_loss   | -0.589   |
|    learning_rate   | 0.0001   |
|    n_updates       | 426134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -57.2    |
| time/              |          |
|    episodes        | 436      |
|    fps             | 18       |
|    time_elapsed    | 23232    |
|    total_timesteps | 435135   |
| train/             |          |
|    actor_loss      | -86.1    |
|    critic_loss     | 619      |
|    ent_coef        | 0.05     |
|    ent_coef_loss   | -0.805   |
|    learning_rate   | 0.0001   |
|    n_updates       | 430134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -61.2    |
| time/              |          |
|    episodes        | 440      |
|    fps             | 18       |
|    time_elapsed    | 23437    |
|    total_timesteps | 439135   |
| train/             |          |
|    actor_loss      | -124     |
|    critic_loss     | 420      |
|    ent_coef        | 0.0556   |
|    ent_coef_loss   | 0.813    |
|    learning_rate   | 0.0001   |
|    n_updates       | 434134   |
----------------------------------
Eval num_timesteps=440000, episode_reward=-80.17 +/- 10.75
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
```

```
| eval/               |          |
|    mean_ep_length   | 1e+03    |
|    mean_reward      | -80.2    |
| time/               |          |
|    total_timesteps  | 440000   |
| train/              |          |
|    actor_loss       | -97      |
|    critic_loss      | 167      |
|    ent_coef         | 0.0592   |
|    ent_coef_loss    | -0.212   |
|    learning_rate    | 0.0001   |
|    n_updates        | 434999   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -63.4    |
| time/               |          |
|    episodes         | 444      |
|    fps              | 18       |
|    time_elapsed     | 23706    |
|    total_timesteps  | 443135   |
| train/              |          |
|    actor_loss       | -104     |
|    critic_loss      | 600      |
|    ent_coef         | 0.069    |
|    ent_coef_loss    | 0.213    |
|    learning_rate    | 0.0001   |
|    n_updates        | 438134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -66.3    |
| time/               |          |
|    episodes         | 448      |
|    fps              | 18       |
|    time_elapsed     | 23910    |
|    total_timesteps  | 447135   |
| train/              |          |
|    actor_loss       | -114     |
|    critic_loss      | 329      |
|    ent_coef         | 0.076    |
|    ent_coef_loss    | 0.0772   |
|    learning_rate    | 0.0001   |
|    n_updates        | 442134   |
----------------------------------
----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -69.5    |
| time/               |          |
|    episodes         | 452      |
|    fps              | 18       |
|    time_elapsed     | 24114    |
|    total_timesteps  | 451135   |
| train/              |          |
|    actor_loss       | -203     |
|    critic_loss      | 1.18e+03 |
|    ent_coef         | 0.0882   |
|    ent_coef_loss    | 0.381    |
|    learning_rate    | 0.0001   |
|    n_updates        | 446134   |
----------------------------------
----------------------------------
```

```
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -69.5    |
| time/              |          |
|    episodes        | 456      |
|    fps             | 18       |
|    time_elapsed    | 24317    |
|    total_timesteps | 455135   |
| train/             |          |
|    actor_loss      | -217     |
|    critic_loss     | 636      |
|    ent_coef        | 0.109    |
|    ent_coef_loss   | -0.314   |
|    learning_rate   | 0.0001   |
|    n_updates       | 450134   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -70.1    |
| time/              |          |
|    episodes        | 460      |
|    fps             | 18       |
|    time_elapsed    | 24520    |
|    total_timesteps | 459135   |
| train/             |          |
|    actor_loss      | -168     |
|    critic_loss     | 1.43e+03 |
|    ent_coef        | 0.117    |
|    ent_coef_loss   | -0.317   |
|    learning_rate   | 0.0001   |
|    n_updates       | 454134   |
----------------------------------
Eval num_timesteps=460000, episode_reward=-82.05 +/- 12.66
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |          |
|    mean_ep_length  | 1e+03    |
|    mean_reward     | -82.1    |
| time/              |          |
|    total_timesteps | 460000   |
| train/             |          |
|    actor_loss      | -172     |
|    critic_loss     | 670      |
|    ent_coef        | 0.115    |
|    ent_coef_loss   | -0.368   |
|    learning_rate   | 0.0001   |
|    n_updates       | 454999   |
----------------------------------
----------------------------------
| rollout/           |          |
|    ep_len_mean     | 1e+03    |
|    ep_rew_mean     | -70.3    |
| time/              |          |
|    episodes        | 464      |
|    fps             | 18       |
|    time_elapsed    | 24789    |
|    total_timesteps | 463135   |
| train/             |          |
|    actor_loss      | -195     |
|    critic_loss     | 1.2e+03  |
|    ent_coef        | 0.108    |
|    ent_coef_loss   | 0.384    |
|    learning_rate   | 0.0001   |
```

```
|    n_updates        | 458134   |
-----------------------------------
-----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -71.9    |
| time/               |          |
|    episodes         | 468      |
|    fps              | 18       |
|    time_elapsed     | 24994    |
|    total_timesteps  | 467135   |
| train/              |          |
|    actor_loss       | -373     |
|    critic_loss      | 1.38e+04 |
|    ent_coef         | 0.118    |
|    ent_coef_loss    | 0.207    |
|    learning_rate    | 0.0001   |
|    n_updates        | 462134   |
-----------------------------------
-----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -73.2    |
| time/               |          |
|    episodes         | 472      |
|    fps              | 18       |
|    time_elapsed     | 25199    |
|    total_timesteps  | 471135   |
| train/              |          |
|    actor_loss       | -226     |
|    critic_loss      | 1.6e+03  |
|    ent_coef         | 0.115    |
|    ent_coef_loss    | 0.197    |
|    learning_rate    | 0.0001   |
|    n_updates        | 466134   |
-----------------------------------
-----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -73.6    |
| time/               |          |
|    episodes         | 476      |
|    fps              | 18       |
|    time_elapsed     | 25406    |
|    total_timesteps  | 475135   |
| train/              |          |
|    actor_loss       | -268     |
|    critic_loss      | 2.18e+03 |
|    ent_coef         | 0.122    |
|    ent_coef_loss    | 0.126    |
|    learning_rate    | 0.0001   |
|    n_updates        | 470134   |
-----------------------------------
-----------------------------------
| rollout/            |          |
|    ep_len_mean      | 1e+03    |
|    ep_rew_mean      | -74.4    |
| time/               |          |
|    episodes         | 480      |
|    fps              | 18       |
|    time_elapsed     | 25608    |
|    total_timesteps  | 479135   |
| train/              |          |
|    actor_loss       | -238     |
```

```
|    critic_loss     | 1.52e+03 |
|    ent_coef        | 0.135    |
|    ent_coef_loss   | 0.0812   |
|    learning_rate   | 0.0001   |
|    n_updates       | 474134   |
----------------------------------
Eval num_timesteps=480000, episode_reward=-90.92 +/- 1.49
Episode length: 1000.00 +/- 0.00
----------------------------------
| eval/              |           |
|    mean_ep_length  | 1e+03     |
|    mean_reward     | -90.9     |
| time/              |           |
|    total_timesteps | 480000    |
| train/             |           |
|    actor_loss      | -280      |
|    critic_loss     | 1.09e+03  |
|    ent_coef        | 0.135     |
|    ent_coef_loss   | -0.236    |
|    learning_rate   | 0.0001    |
|    n_updates       | 474999    |
----------------------------------
----------------------------------
| rollout/           |           |
|    ep_len_mean     | 1e+03     |
|    ep_rew_mean     | -75.5     |
| time/              |           |
|    episodes        | 484       |
|    fps             | 18        |
|    time_elapsed    | 25873     |
|    total_timesteps | 483135    |
| train/             |           |
|    actor_loss      | -195      |
|    critic_loss     | 759       |
|    ent_coef        | 0.133     |
|    ent_coef_loss   | 0.277     |
|    learning_rate   | 0.0001    |
|    n_updates       | 478134    |
----------------------------------
----------------------------------
| rollout/           |           |
|    ep_len_mean     | 1e+03     |
|    ep_rew_mean     | -75.7     |
| time/              |           |
|    episodes        | 488       |
|    fps             | 18        |
|    time_elapsed    | 26077     |
|    total_timesteps | 487135    |
| train/             |           |
|    actor_loss      | -286      |
|    critic_loss     | 2.39e+03  |
|    ent_coef        | 0.136     |
|    ent_coef_loss   | -0.0232   |
|    learning_rate   | 0.0001    |
|    n_updates       | 482134    |
----------------------------------
----------------------------------
| rollout/           |           |
|    ep_len_mean     | 1e+03     |
|    ep_rew_mean     | -75.5     |
| time/              |           |
|    episodes        | 492       |
|    fps             | 18        |
|    time_elapsed    | 26280     |
```

```
|    total_timesteps | 491135   |
```

```
# --------------------
# Helpers to eval with true env rewards
# --------------------
def load_eval_env_with_stats():
    env = make_eval_env(SEED + 100)
    env = VecNormalize.load(os.path.join(LOG_DIR, "vecnormalize.pkl"), env)
    env.training = False
    env.norm_reward = False  # report true env rewards
    return env
```

```
|    ep_rew_mean     | 75.8     |
```

```
# --------------------
# Evaluate FINAL model
# --------------------
eval_env_final = load_eval_env_with_stats()
final_model = SAC.load(final_model_path, env=eval_env_final)
mean_r, std_r = evaluate_policy(final_model, eval_env_final, n_eval_episodes=FINAL_EVAL_EPISODES, deterministic=True)
print(f"[FINAL @500k] Mean reward: {mean_r:.2f} +/- {std_r:.2f}")
eval_env_final.close()
```

```
|    learning_rate   | 0.0001   |
|    n_updates       | 490134   |
-----------------------------------
[FINAL @500k] Mean reward: 774.95 +/- 10.34
```

```
# --------------------
# Evaluate BEST model
# --------------------
best_model_path = os.path.join(LOG_DIR, "best_model.zip")
if os.path.exists(best_model_path):
    eval_env_best = load_eval_env_with_stats()
    best_model = SAC.load(best_model_path, env=eval_env_best)
    mean_r_b, std_r_b = evaluate_policy(best_model, eval_env_best, n_eval_episodes=FINAL_EVAL_EPISODES, deterministic=True)
    print(f"[BEST  @500k] Mean reward: {mean_r_b:.2f} +/- {std_r_b:.2f}")
    eval_env_best.close()
else:
    print("No best_model.zip found (EvalCallback may not have improved over initial policy).")
```

```
|    n_updates       | 494134   |
[BEST  @500k] Mean reward: 682.02 +/- 268.29
Eval num_timesteps=500000, episode_reward=-81.52 +/- 7.12
```

```
# --------------------
# Record short video (best if available else final)
# --------------------
def record_video(model_path, out_prefix="best_model"):
    env = load_eval_env_with_stats()
    env = VecVideoRecorder(
        env,
        VIDEO_DIR,
        record_video_trigger=lambda step: step == 0,
        video_length=1000,  # CarRacing horizon ~1000 steps
        name_prefix=f"{out_prefix}_car_racing_sac_500k",
    )
    obs = env.reset()
    # Load model once outside the loop
    policy = SAC.load(model_path, env=env)
    for _ in range(1000):
        action, _ = policy.predict(obs, deterministic=True)
```

```
            obs, rewards, dones, infos = env.step(action)
            if dones[0]:
                break
        env.close()

    if os.path.exists(best_model_path):
        record_video(best_model_path, out_prefix="best_model")
    else:
        record_video(final_model_path, out_prefix="final_model")

    print(f"Video(s) saved to: {VIDEO_DIR}")
```

```
Saving video to /content/videos/best_model_car_racing_sac_500k-step-0-to-step-1000.mp4
/usr/local/lib/python3.12/dist-packages/moviepy/config_defaults.py:47: SyntaxWarning: invalid escape sequence '\P'
  IMAGEMAGICK_BINARY = r"C:\Program Files\ImageMagick-6.8.8-Q16\magick.exe"
Moviepy - Building video /content/videos/best_model_car_racing_sac_500k-step-0-to-step-1000.mp4.
Moviepy - Writing video /content/videos/best_model_car_racing_sac_500k-step-0-to-step-1000.mp4

/usr/local/lib/python3.12/dist-packages/jupyter_client/session.py:203: DeprecationWarning: datetime.datetime.utcnow() is deprecated and scheduled for removal ir
  return datetime.utcnow().replace(tzinfo=utc)
Moviepy - Done !
Moviepy - video ready /content/videos/best_model_car_racing_sac_500k-step-0-to-step-1000.mp4
Video(s) saved to: ./videos
```

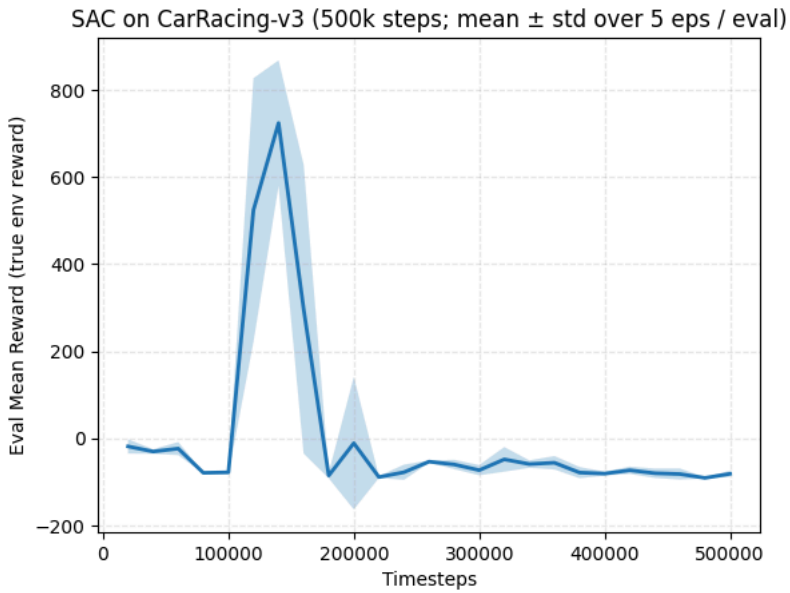
```
# See it
import glob
glob.glob("/content/videos/*.mp4")

# Download to your computer
from google.colab import files
files.download("/content/videos/best_model_car_racing_sac_500k-step-0-to-step-1000.mp4")
```

```
from google.colab import drive
drive.mount('/content/drive')

import os, shutil
DEST = "/content/drive/MyDrive/CarRacingVideos"
os.makedirs(DEST, exist_ok=True)
shutil.copy("/content/videos/best_model_car_racing_sac_500k-step-0-to-step-1000.mp4",
            f"{DEST}/best_model_car_racing_sac_500k-step-0-to-step-1000.mp4")
print("Copied to:", DEST)
```

```
Mounted at /content/drive
Copied to: /content/drive/MyDrive/CarRacingVideos
```

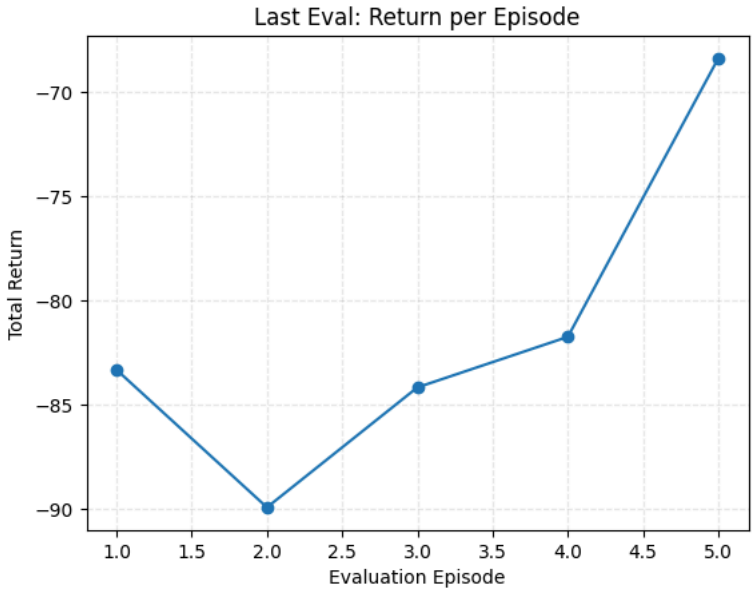
```
# ---------------------
# Plot evaluation curve (from EvalCallback)
# ---------------------
eval_npz = os.path.join(LOG_DIR, "evaluations.npz")
if os.path.exists(eval_npz):
    data = np.load(eval_npz)
    timesteps = data["timesteps"]
    results = data["results"]
```

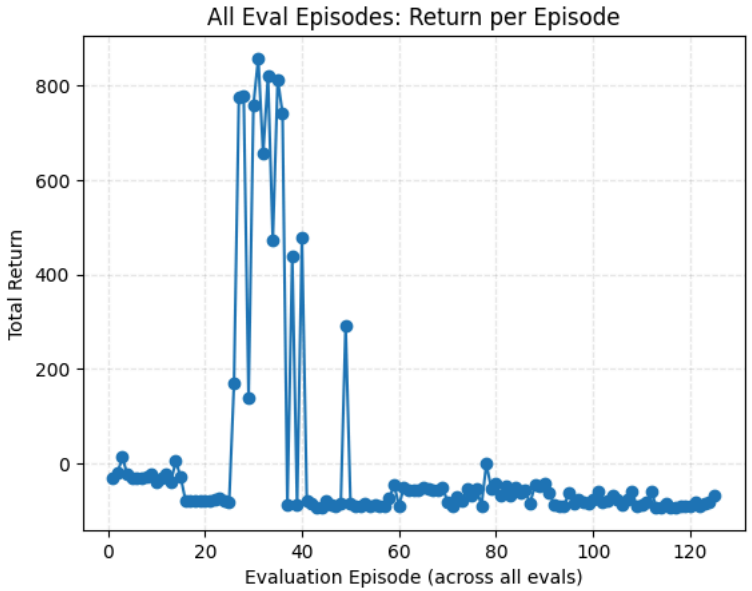
```
    mean_results = results.mean(axis=1)
    std_results = results.std(axis=1)

    plt.figure()
    plt.plot(timesteps, mean_results, linewidth=2)
    plt.fill_between(timesteps, mean_results - std_results, mean_results + std_results, alpha=0.25)
    plt.xlabel("Timesteps")
    plt.ylabel("Eval Mean Reward (true env reward)")
    plt.title(f"SAC on {ENV_ID} (500k steps; mean ± std over {EVAL_EPISODES} eps / eval)")
    plt.grid(True, linestyle="--", alpha=0.3)
    plt.show()
else:
    print("evaluations.npz not found; was EvalCallback attached?")
```




```
import numpy as np
import matplotlib.pyplot as plt
import os

LOG_DIR = "./logs/CarRacing-v3"  # change if yours differs
npz_path = os.path.join(LOG_DIR, "evaluations.npz")
data = np.load(npz_path)

# results shape: (num_evals, episodes_per_eval)
results = data["results"]

# Option 1: last evaluation only
last_returns = results[-1]  # shape (episodes_per_eval,)
plt.figure()
plt.plot(range(1, len(last_returns)+1), last_returns, marker="o")
plt.xlabel("Evaluation Episode")
plt.ylabel("Total Return")
plt.title("Last Eval: Return per Episode")
```

```
plt.grid(True, linestyle="--", alpha=0.3)
plt.show()

# Option 2: all evaluations flattened into a single sequence
flat_returns = results.reshape(-1)
plt.figure()
plt.plot(range(1, len(flat_returns)+1), flat_returns, marker="o")
plt.xlabel("Evaluation Episode (across all evals)")
plt.ylabel("Total Return")
plt.title("All Eval Episodes: Return per Episode")
plt.grid(True, linestyle="--", alpha=0.3)
plt.show()
```






```
# Rebuild eval env with saved VecNormalize stats, then disable reward norm
eval_env = load_eval_env_with_stats()

# Load whichever model you want to plot
model_to_plot = SAC.load(best_model_path, env=eval_env)  # or final_model_path

# Get per-episode returns
episode_returns, episode_lengths = evaluate_policy(
```

```
        model_to_plot, eval_env,
        n_eval_episodes=20,
        deterministic=True,
        return_episode_rewards=True
    )

    import matplotlib.pyplot as plt
```