Trainer

weights
θ

trajectory batches

Runner

Envs
π

- Recieves Minibatches from the trainer
- Implements a learning algorithm (PPO)
- Returns a policy parametrizations (weights) to the runner

- Wraps one or more physics simulation to generate trajectories
- Concatenate the trajectories into minibatches for the trainer