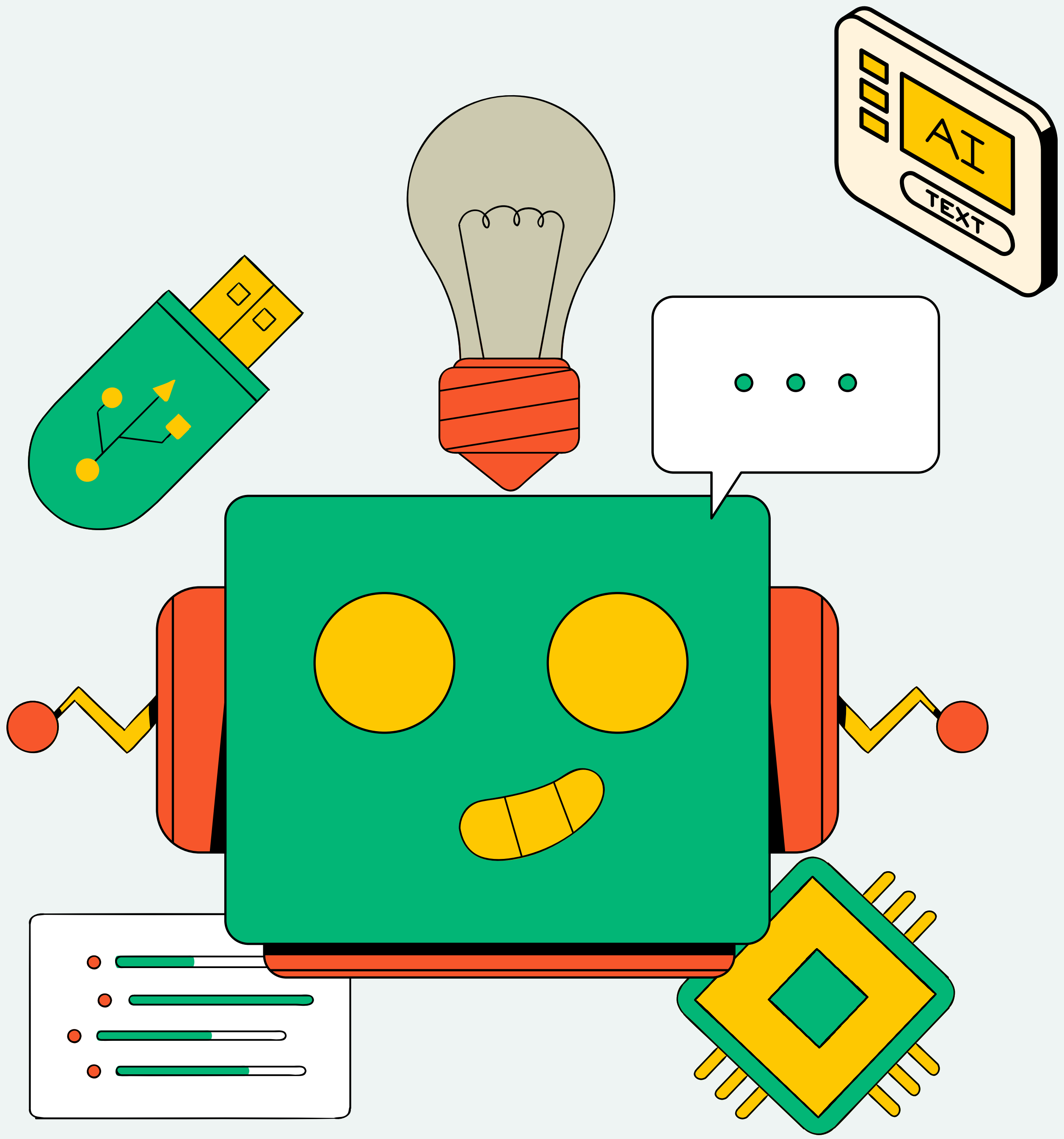
ROBOT LEARNING

2024-2025

DOMAIN RANDOMIZATION IN THE GYM ENVIROMENT

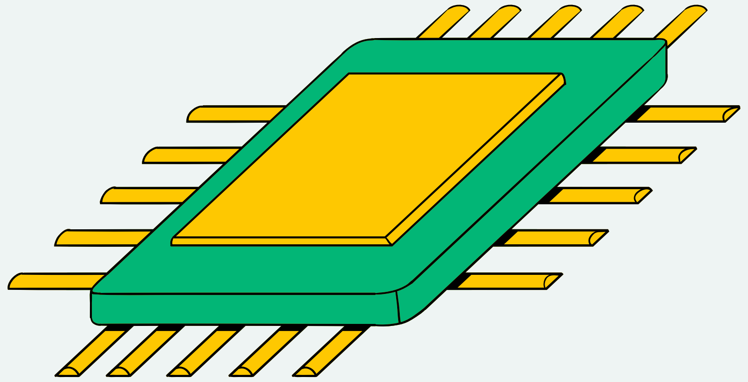


PRESENTED BY:

MAGGIULLI CLAUDIA
S332252

PENZA ALESSANDRA
S331062

PORCELLI FRANCESCA
S324804



KEY CONCEPTS

Uniform Domain Randomization

Sim-to-Real

Adaptive Domain Randomization

Hopper Envioment

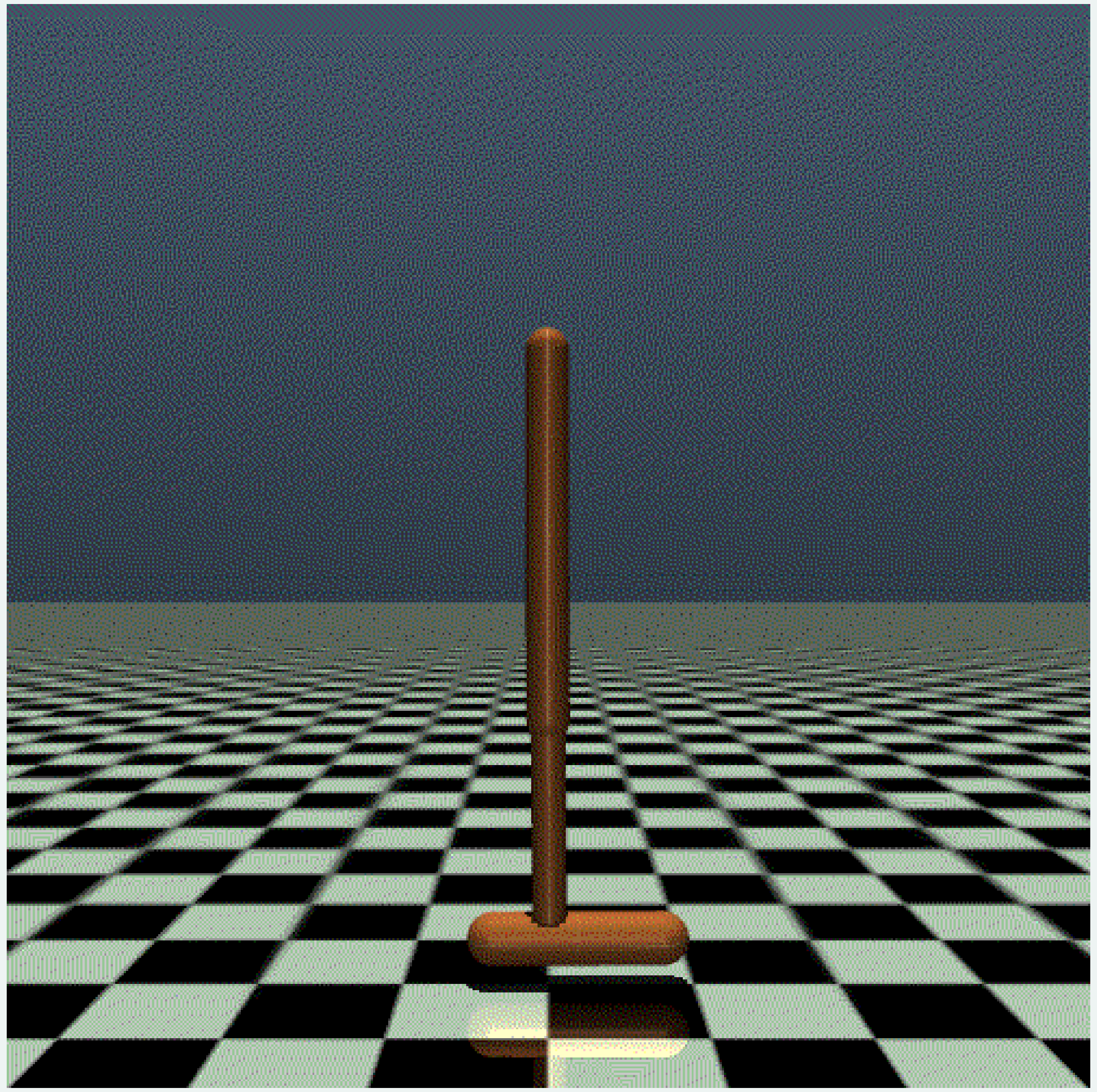
Half Cheetah Enviroment

Experiments

HOPPER ENVIRONMENT

2D, single-legged robot with four parts: torso, thigh, leg, and foot.

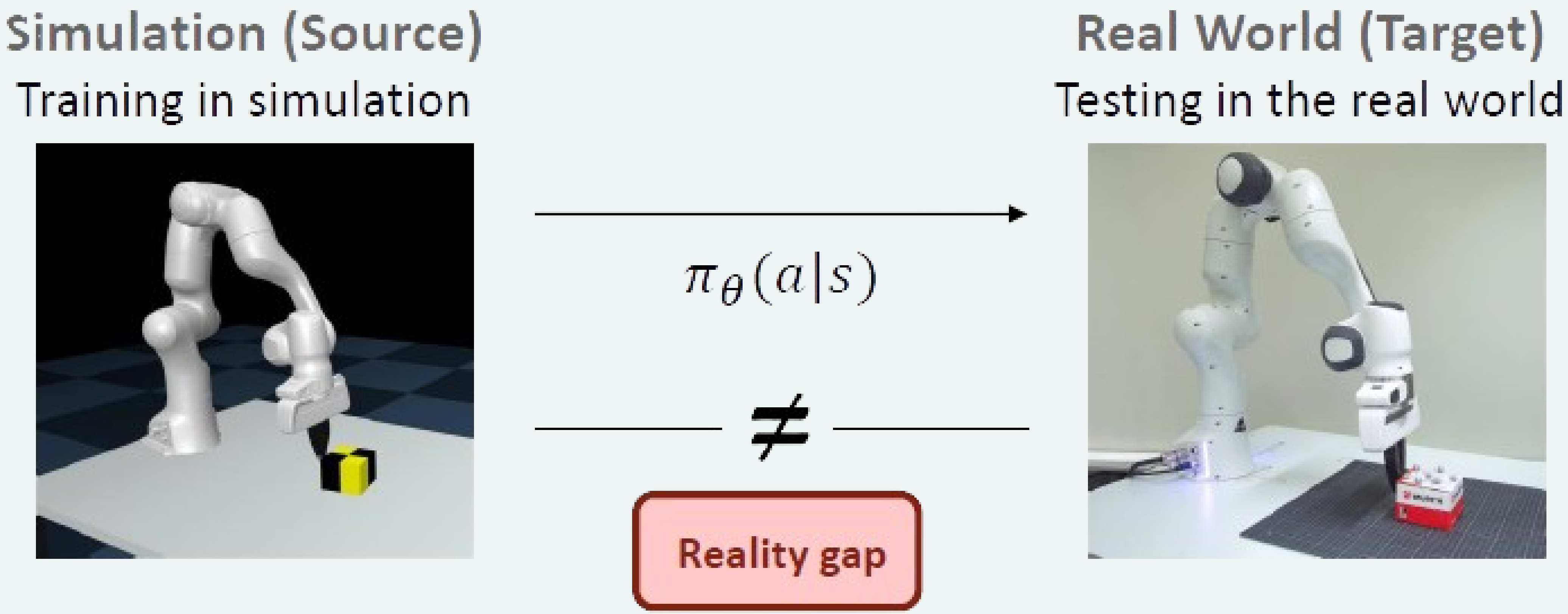
Goal: jumping efficiently while maximizing horizontal speed.



State space: 11-dimensional vector for dynamics (joint angles, linear velocity and angular velocity) → continuous space
Action space: Box(-1, 1, (3,), float32) for each joint's applied torques (thigh, leg, foot) → continuous space

SIM-TO-REAL TRANSFER

Sim-to-Real Transfer Problem:
Reality gap ⇒ policies trained in simulation struggle to generalize to real-world scenarios



Solution: **Sim-to-Sim setup**
Introducing variations between a **source** and a **target** environments to simulate transfer challenges.

TRAINING PIPELINE: IMPLEMENTED ALGORITHM

Proximal Policy Optimization (PPO):
implements clipped objective function (prevent large policy updates).

Soft Actor-Critic (SAC):
maximizes both reward and entropy (diverse actions for better exploration).

WHY PPO FOR OUR PROJECT?

PPO → well-suited for structured environments with predictable dynamics
SAC → better in highly randomized and complex settings.

Evaluation of policy transfer from a source domain (lighter torso) to a target domain.

EXPERIMENTAL RESULTS ON HOPPER ENVIRONMENT

CONFIGURATION	AVERAGE RETURN ± STD
source → source	867.74 ± 263.64
source → target	932.24 ± 108.21
target → target	1595.30 ± 31.42

Best performance:
target → target without the reality gap.

source → target lower results: no direct experience with the target domain during training
target → target not best option: high costs and safety concerns