Learning Robust Bipedal Locomotion Across Variable Farm Soil Structures via Multistages Deep Reinforcement Learning

Anhar Risnumawan (D2) · Kubota's Lab · Tokyo Metropolitan University · **P77** · **CONFIDENTIAL**

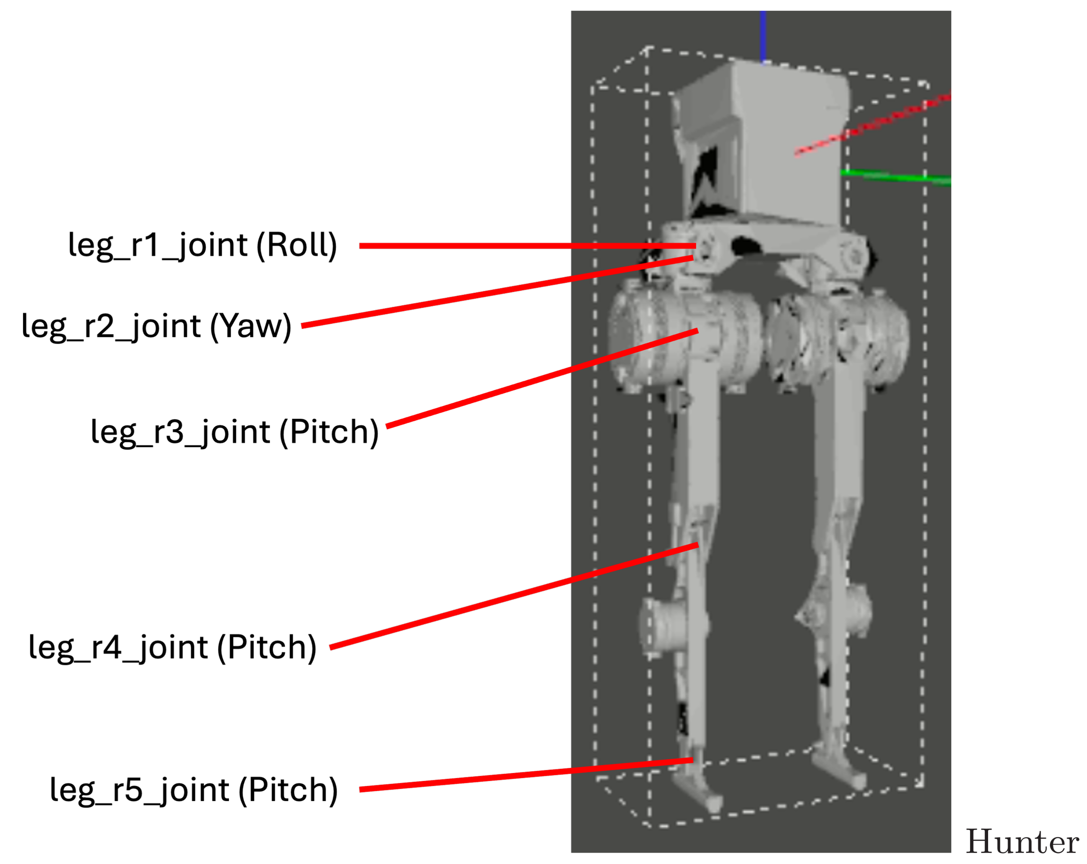
TOKYO METROPOLITAN UNIVERSITY

1. Introduction

Agricultural robotics faces significant challenges when operating in **unstructured farm environments**. Traditional bipedal robots struggle with variable soil conditions, **ranging from hard-packed earth to soft, wet to dry terrain**. This research presents a novel multistages approach using deep reinforcement learning to train a bipedal robot (Hunter) for robust locomotion across agricultural soil structures. Our three-stage curriculum progressively increases terrain complexity: plane terrain → soft soil → farm soil, enabling the robot to **develop adaptive walking strategies essential for agricultural applications**.

2. Robot Platform

EC-Hunter80-V01 Specifications:
- **DOF:** 10 (legs only)
- **Mass:** 12.7 kg
- **Height:** 0.6 m standing
- **Structure:** Bipedal with base platform
- **Actuators:** 23.7 Nm max torque



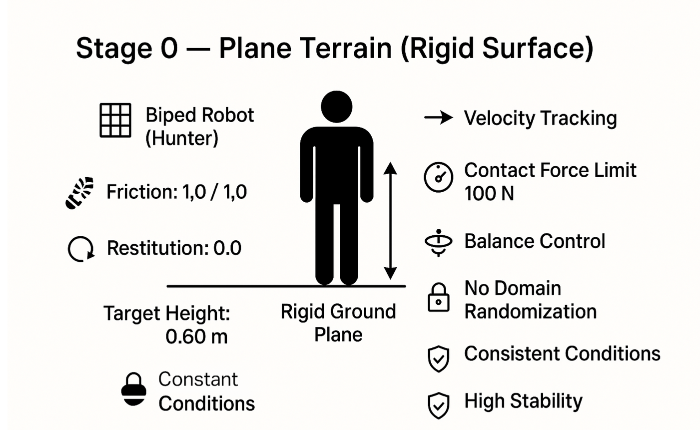
Robot Joint Configuration

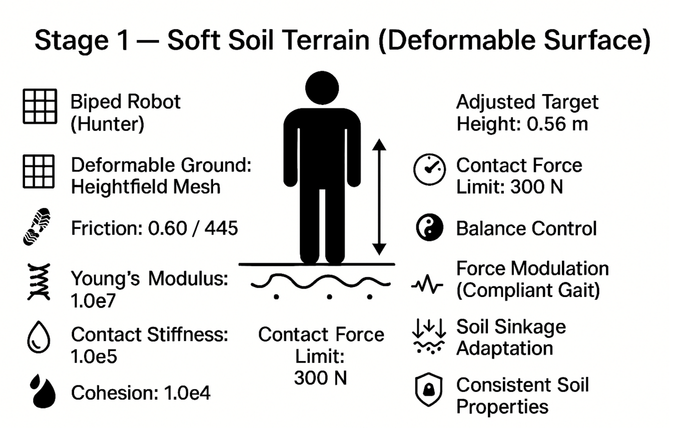
Joint Mapping:
- Hip: Roll, Yaw, Pitch (3 DOF/leg)
- Knee: Flexion (1 DOF/leg)
- Ankle: Dorsiflexion (1 DOF/leg)

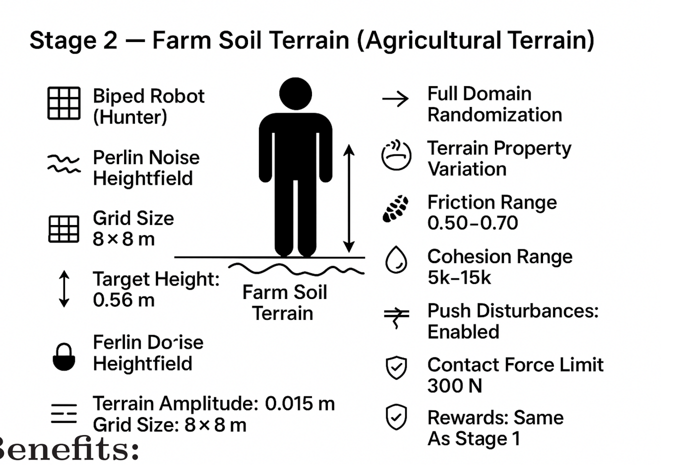
3. Contribution: Three-Stage Curriculum Learning

Our approach progressively increases environmental complexity to develop robust locomotion skills:

Stage	Environment	Terrain Properties	Domain Randomization	Reward Function	Skills Developed	Training Purpose
Stage 0	Plane Terrain (Rigid surface)	• Mesh: plane • Friction: 1.0/1.0 • Restitution: 0.0 • High stability	None • Push disturbances • No parameter variation • Consistent conditions	Standard locomotion rewards • Target height: 0.6m • Contact force limit: 100N	• Basic bipedal gait • Joint coordination • Velocity tracking • Balance control	Foundation stage for loading pretrained models • Establishes baseline walking patterns
Stage 1	Soft Soil (Deformable surface)	• Mesh: heightfield • Young's modulus: 1.0e7 • Contact stiffness: 1.0e5 • Cohesion: 1.0e4 • Friction: 0.60/0.45	None • No randomization • Consistent soil properties • Stable learning conditions	Soil-adapted rewards • Adjusted height: 0.56m • Contact force limit: 300N • Reduced penalties for slip/orientation	• Adaptive foot placement • Force modulation • Compliant surface gait • Soil sinkage adaptation	Introduces soil physics without complexity of randomization • Smooth progression from rigid to soft
Stage 2	Farm Soil (Agricultural terrain)	• Same soil physics • Terrain amplitude: 0.015 • Perlin noise generation • 8x8m terrain grid • 64x64 resolution	Full randomization • Push disturbances: disabled • Terrain property variation • Friction range: [0.50,0.70] • Cohesion range: [5k,15k]	Same as Stage 1 • Handles variability • Maintains tracking performance under uncertainty	• Robust terrain adaptation • Multi-modal locomotion • Environmental adaptation • Disturbance rejection • Parameter robustness	Maximum robustness training with all environmental variations • Prepares for real-world deployment







Our contributions:
- Introduce a three-stage soil-aware curriculum (rigid → soft soil → farm soil)
- Design soil-tuned rewards and PD gains
- Demonstrate robust biped locomotion in photorealistic Isaac Sim farm-soil with push recovery-targeting agricultural robots

Progressive Learning Benefits:
- Stable foundation building in Stage 0
- Gradual complexity introduction prevents training instabilities
- Transfer learning between stages accelerates convergence
- Final policy generalizes across diverse soil conditions

4. Deep RL Architecture

Algorithm: Proximal Policy Optimization (PPO)
Neural Network:
- Actor: 512→256→128 (ELU activation)
- Critic: 512→256→128 (ELU activation)
- Output: 10D action space (joint commands)

Key Reward Components:
- **Tracking:** Linear/angular velocity following
- **Stability:** Orientation and height maintenance
- **Energy:** Torque and acceleration penalties
- **Contact:** Proper foot placement rewards
- **Safety:** Joint limits and collision avoidance

Observation Space (45D):
- Base orientation, velocity
- Joint positions, velocities
- Contact states
- Command targets

5. Technical Innovations

Reward System Design:
- Multi-objective optimization balancing performance and efficiency
- Hierarchical penalty structure prioritizing safety
- Contact-aware rewards for foot-ground interaction

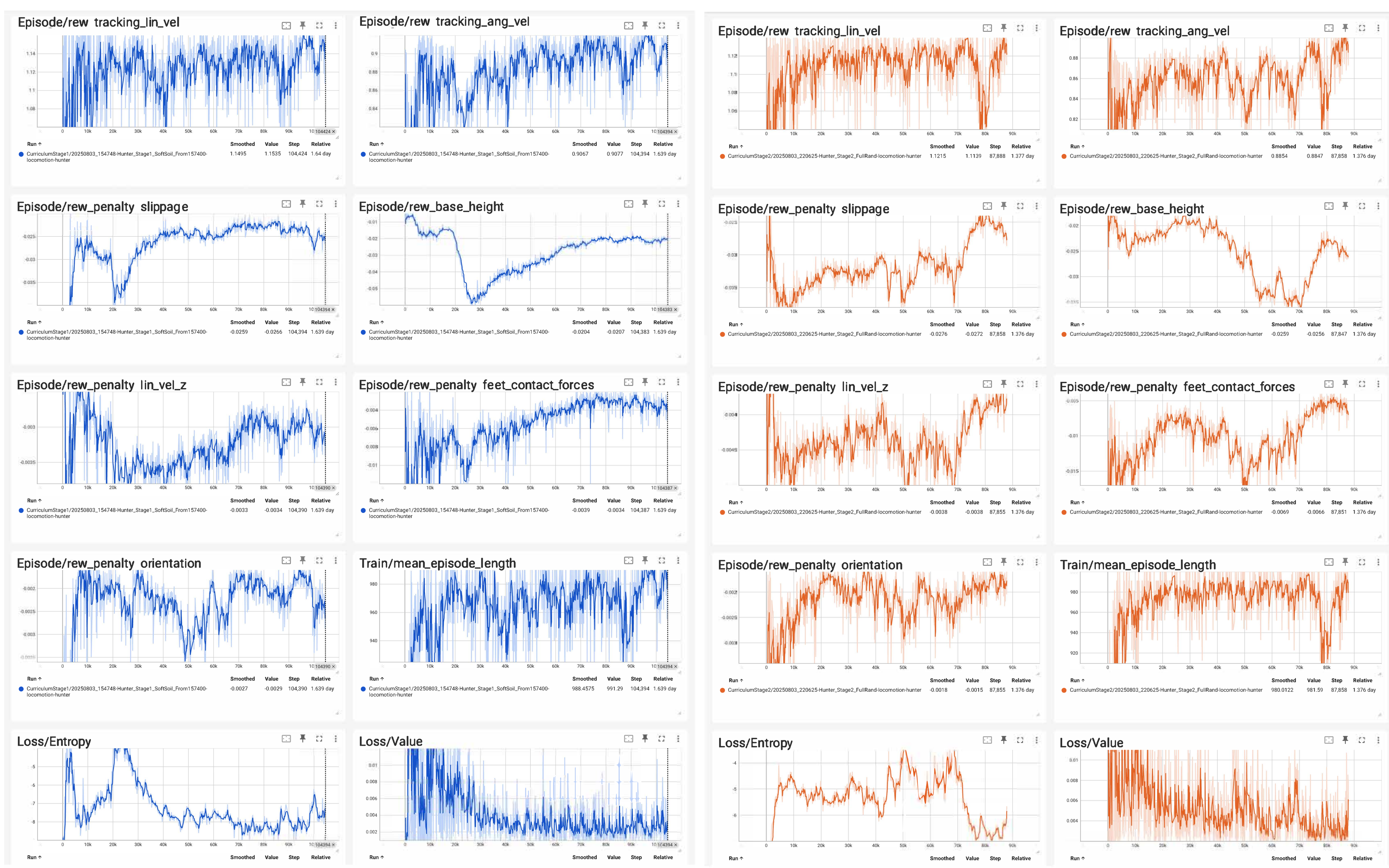
Morphology Adaptation:
- URDF standardization for policy transfer
- Joint limit calibration matching hardware specs
- Contact sensor mapping for ground interaction

Training Optimizations:
- Episode timing: 50 Hz control frequency
- Batch learning: 5 epochs per iteration
- Curriculum scheduling: Progressive difficulty increase

6. Related Work

- **Sim-to-real legged locomotion.** Learning-based controllers achieve robust quadruped locomotion in natural terrain and drastically reduced training times via massive parallelism [1,2]. Rapid Motor Adaptation handles unseen dynamics through online adaptation [3].
- **Deformable terrain.** High-speed locomotion on soft sand using contact-aware training highlights the need to model soil compliance [4].
- **Imitation and skill priors.** Motion imitation (e.g., DeepMimic) seeds policies with natural gaits that transfer well [5].
- **Hybrid control.** Combining RL with heuristic components can improve biped stability and sim-to-real transfer [6].
- **Curriculum learning.** Structured curricula enable reliable biped learning over complex terrains [7].
- **Humanoids on compliant ground.** Recent results demonstrate robust DRL walking on HRP-5P over uneven, compliant terrain [8].

7. Experiments

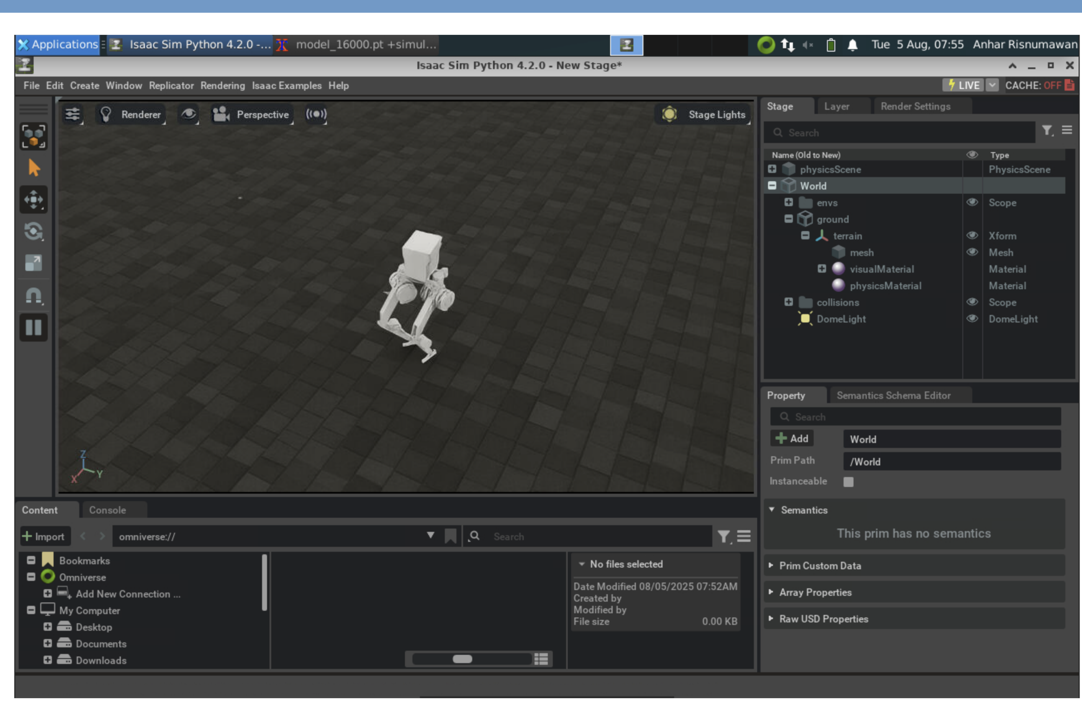


Learning curves (left is stage 1, right is stage 2) showing rewards convergence.

Key Training Achievements:
- Successful locomotion learning within $\approx 100K$ iterations
- Stable upright posture maintenance (base_height penalty minimized)
- Effective velocity tracking (tracking rewards maximized)
- Energy-efficient gait patterns (torque penalties reduced)
- Robust balance control under external disturbances (0.5 m/s push resistance)

Performance Metrics:
- Walking speed: ±0.5 m/s target tracking
- Episode length: 20 seconds (1000 control steps at 50 Hz)
- Success rate: >90% episode completion
- Balance recovery: Maintains stability under lateral pushes

8. Soil Adaptation Capabilities



Hunter robot navigating across different soil environments: plane terrain → soft soil → farm soil

Stage 2 - Farm Soil Locomotion:
- Integration of plane terrain skills with soft soil adaptations
- Robust performance across wet to dry agricultural soil conditions
- Possible applicability for farming operations and crop monitoring

9. Conclusions & Future Work

Conclusions:
1. Demonstrated multistages learning for agricultural soil locomotion
2. Successful deep RL training of Hunter biped
3. Comprehensive reward system for bipedal stability

Future Directions:
- Real hardware deployment and validation
- Long-duration autonomous field operations

10. References

1. Lee, J., et al. "Learning quadrupedal locomotion over challenging terrain." Science robotics 5.47 (2020).
2. Rudin, N., et al. "Learning to walk in minutes using massively parallel deep reinforcement learning." arXiv:2109.11978 (2022).
3. Kumar, A., et al. "RMA: Rapid Motor Adaptation for Legged Robots." RSS (2021).
4. Choi, S., et al. "Learning quadrupedal locomotion over deformable terrain." Science Robotics 8.74 (2023).
5. Peng, X. B., et al. "DeepMimic: Example-guided deep reinforcement learning of physics-based character skills." ACM TOG 37.4 (2018).
6. Wang, Z., et al. "Hybrid locomotion control of biped robots based on reinforcement learning." IEEE TSMC 53.1 (2022).
7. Tidd, B., et al. "Guided curriculum model-based reinforcement learning for walking on uneven terrain." IROS (2020).
8. Singh, R., et al. "Robust humanoid locomotion using deep reinforcement learning with compliant contact models." arXiv:2501.00123 (2025).