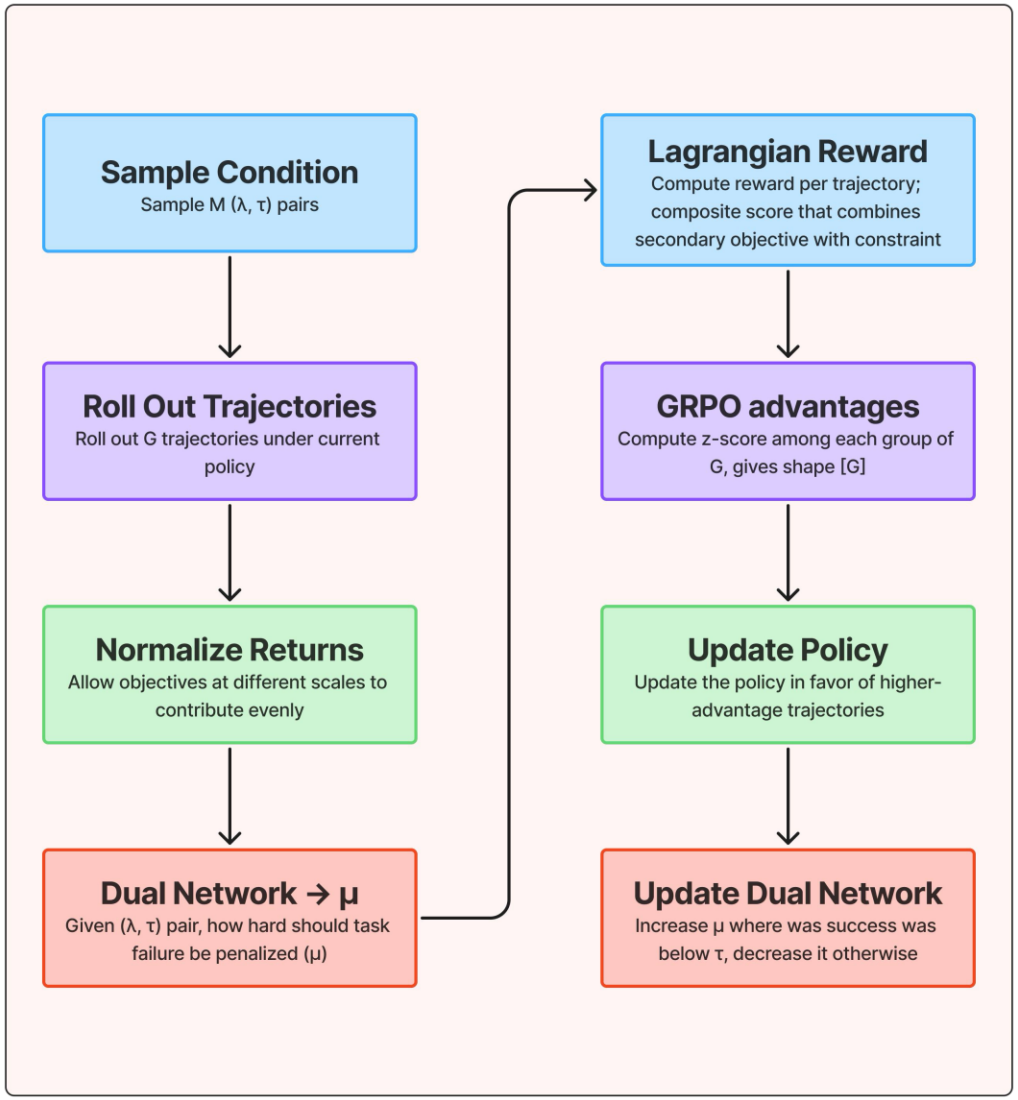
WEEK 4 UPDATE

CODY HILL • 6/15/26

Problem Overview



Task

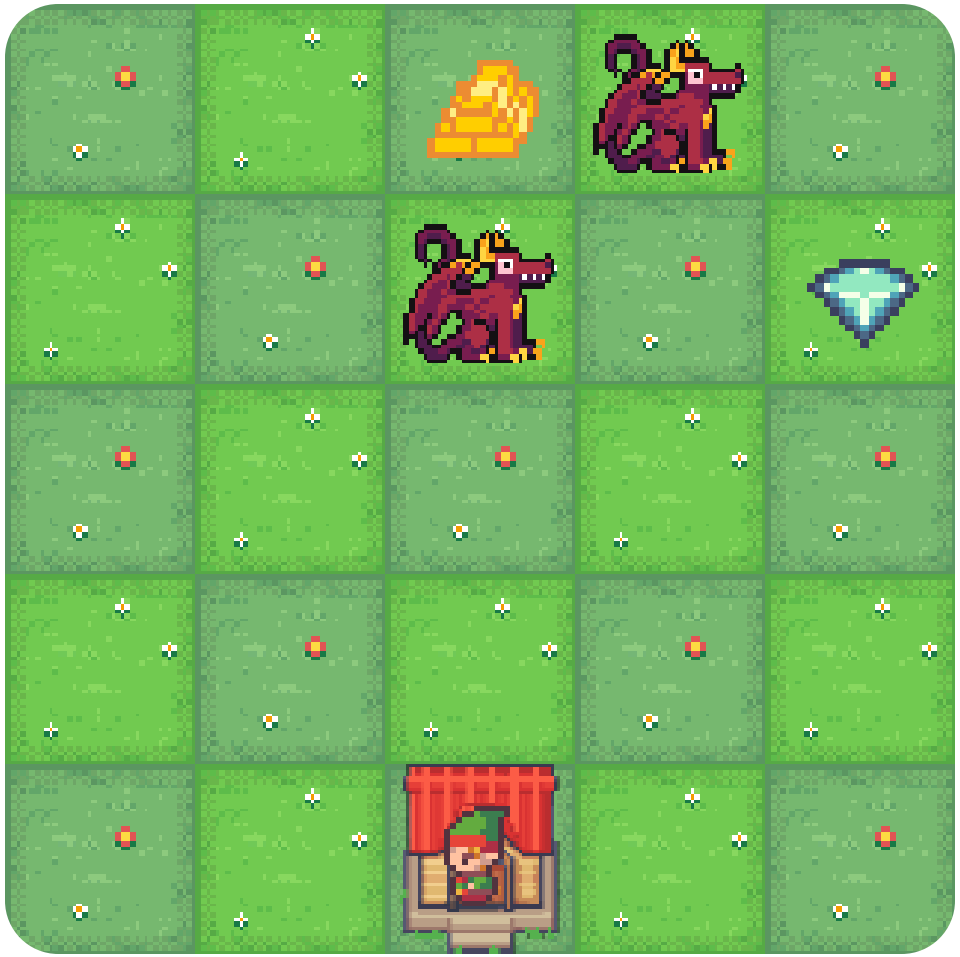
- GRPO with task-success floor
- Optimized secondary objectives
- Policy π with maximum returns

Problem Overview

Success*	0.64	0.85	0.85
Efficiency	0.84	0.41	0.77
Safety	0.73	0.51	0.68
Smoothness	0.75	0.43	0.75
$\tau = 0.8$	Bad	Feasible	Optimal

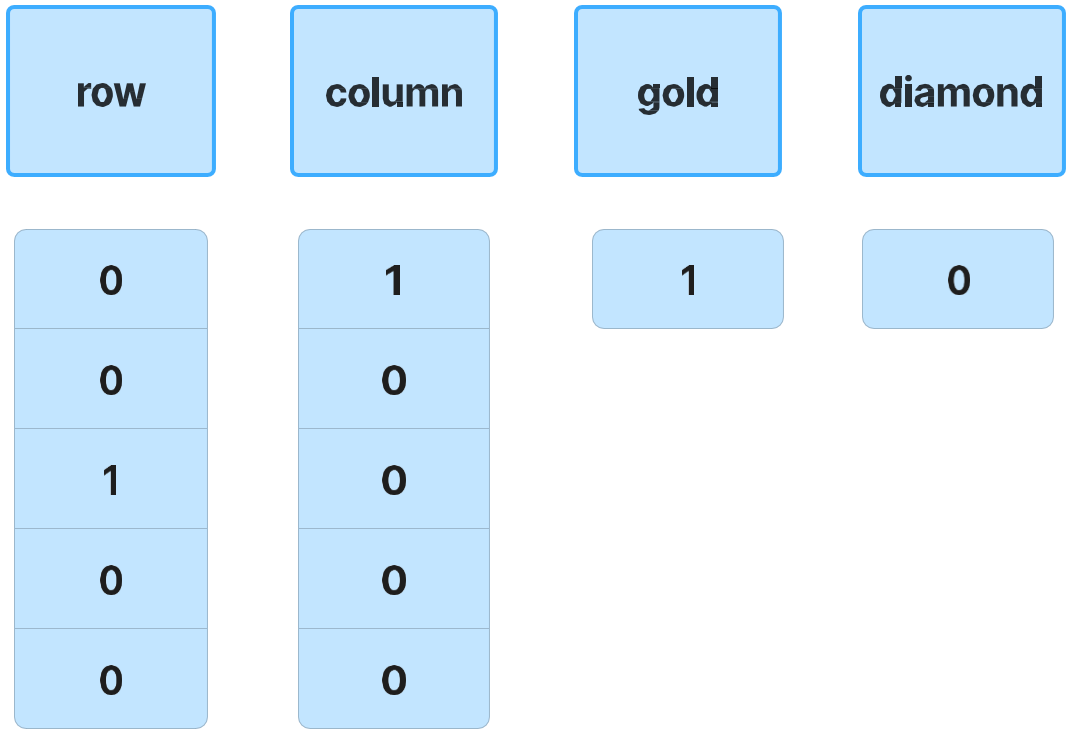
Progress: New Environment

- “Resource-Gathering”
- 3 Objectives
  - Survival (Constrained)
  - Gold
  - Diamond
- Known Pareto Front
- No continuous tradeoff



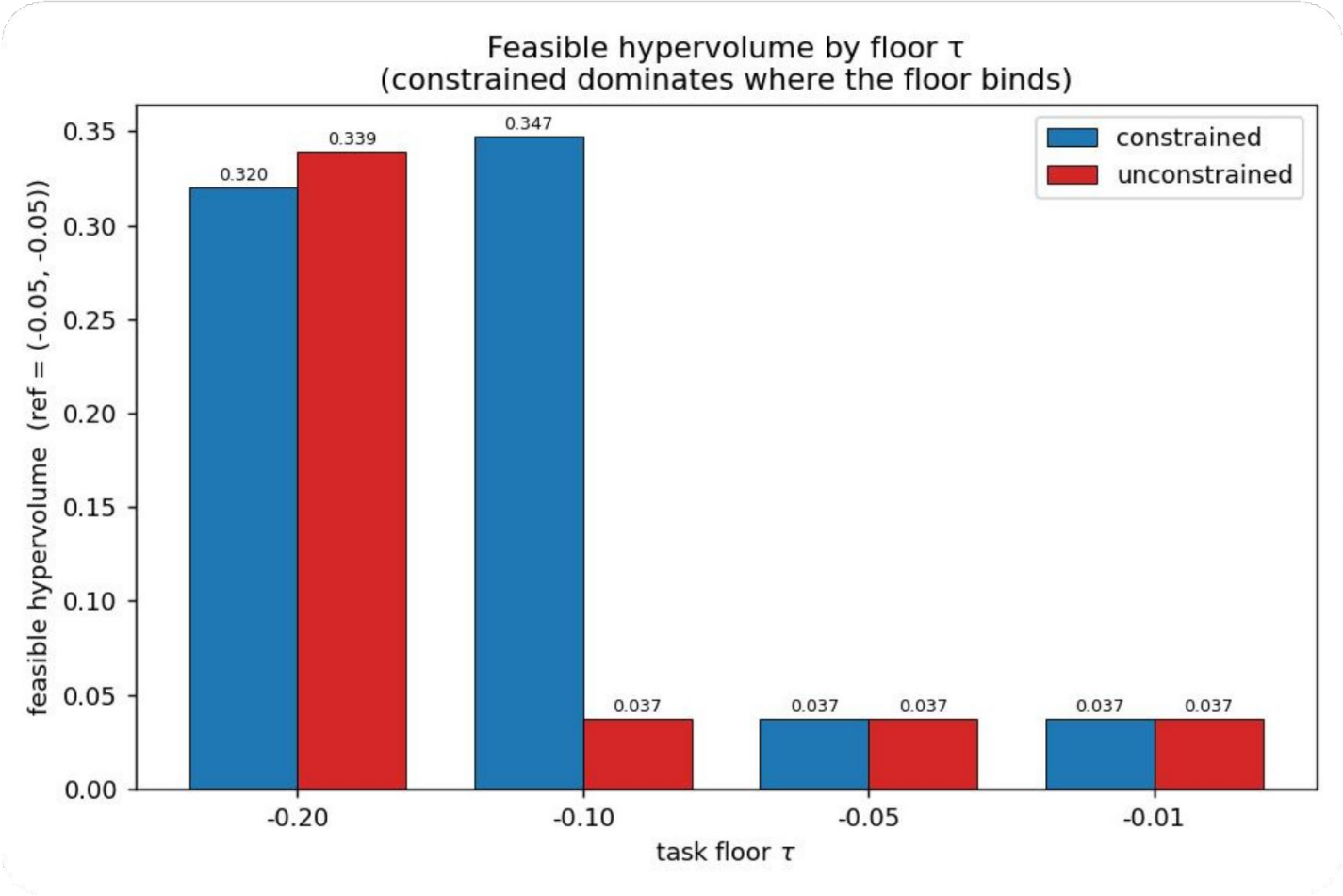
Progress: Featurization

- Helps policy learn faster
- One-hot encoding
- [col, row, gold, diamond]
- 4 dim. input → 12 dim. input
- Hot-row, hot-column



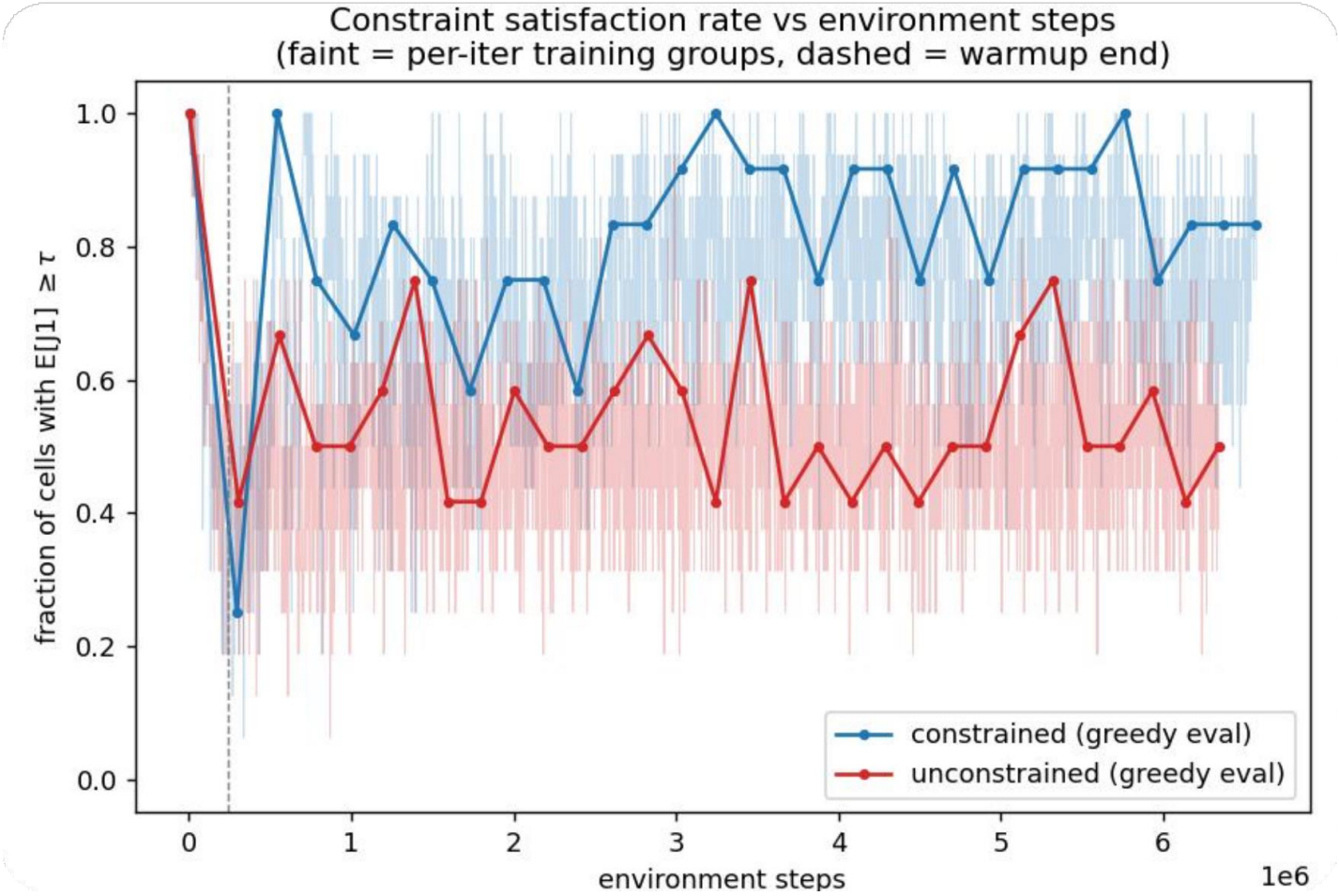
Progress: Hypervolume

- Recovered area
- Expected results



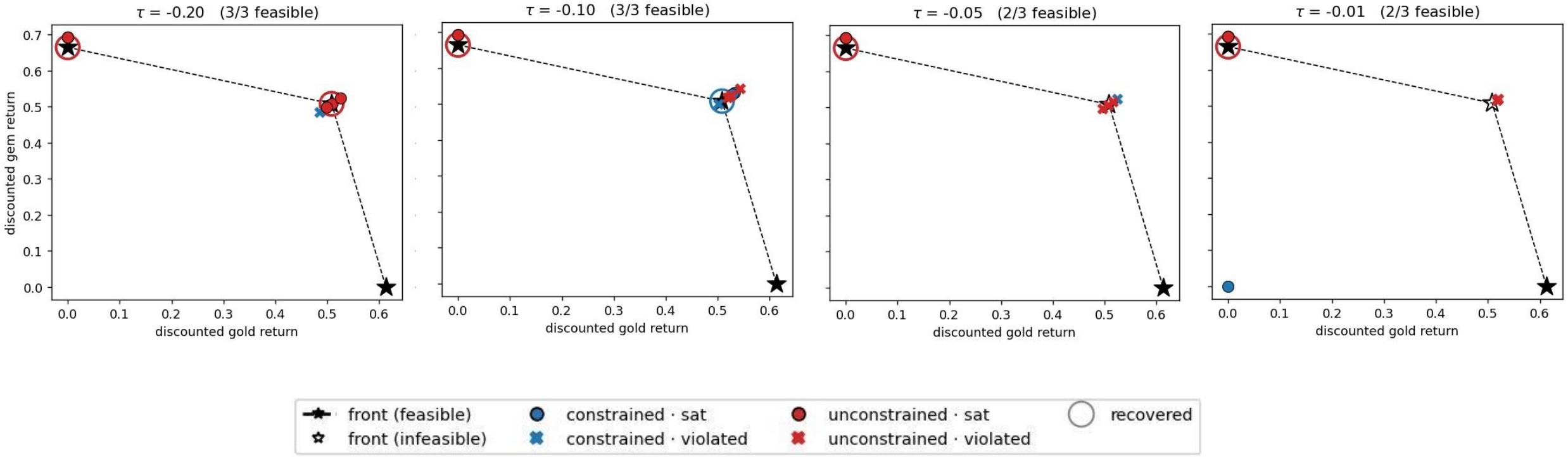
Progress: Constraint Satisfaction

- (λ, τ) pairs
- Satisfy τ floor
- Expected behavior
- Minor instability



Progress: Pareto Front

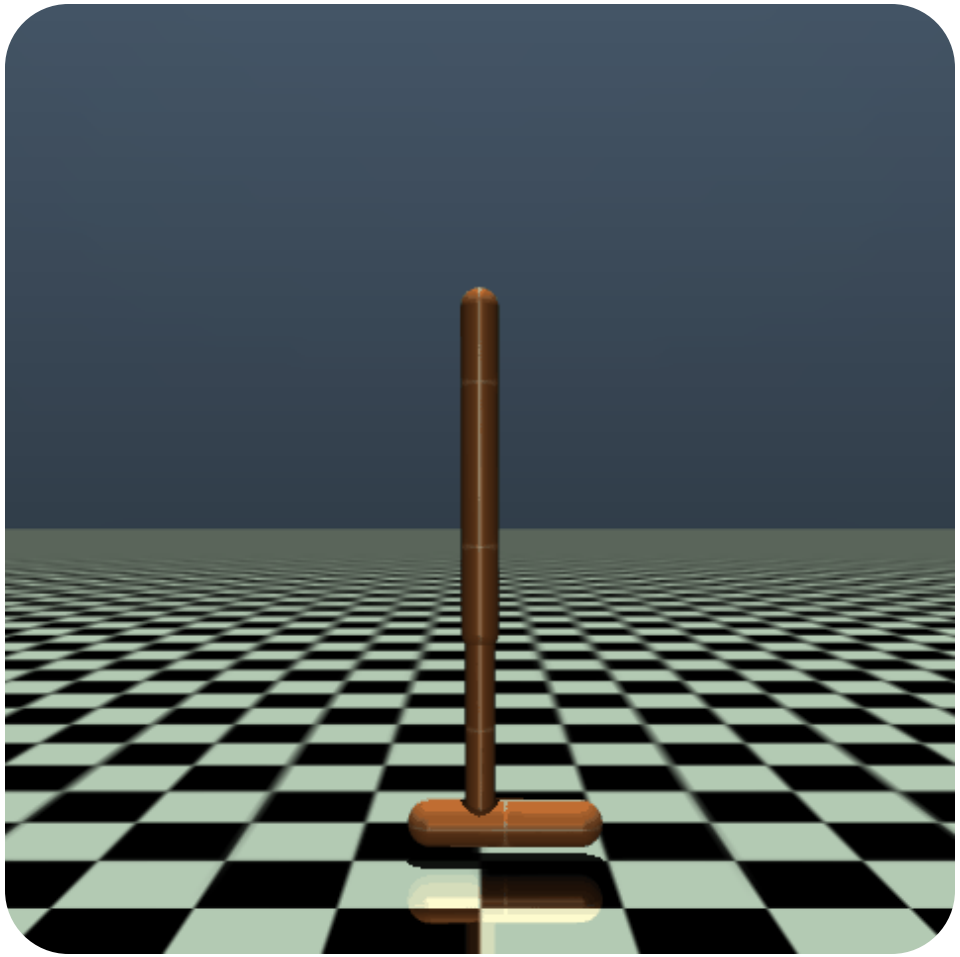
Pareto Recovery by floor τ. (O = satisfied, X = violated, ring = feasibly recovered)



Current Work and Next Steps

Next Steps

- Continuous Environment
- Finish in coming days
- VLA after confirmation



WEEK 4 UPDATE

CODY HILL • 6/15/26