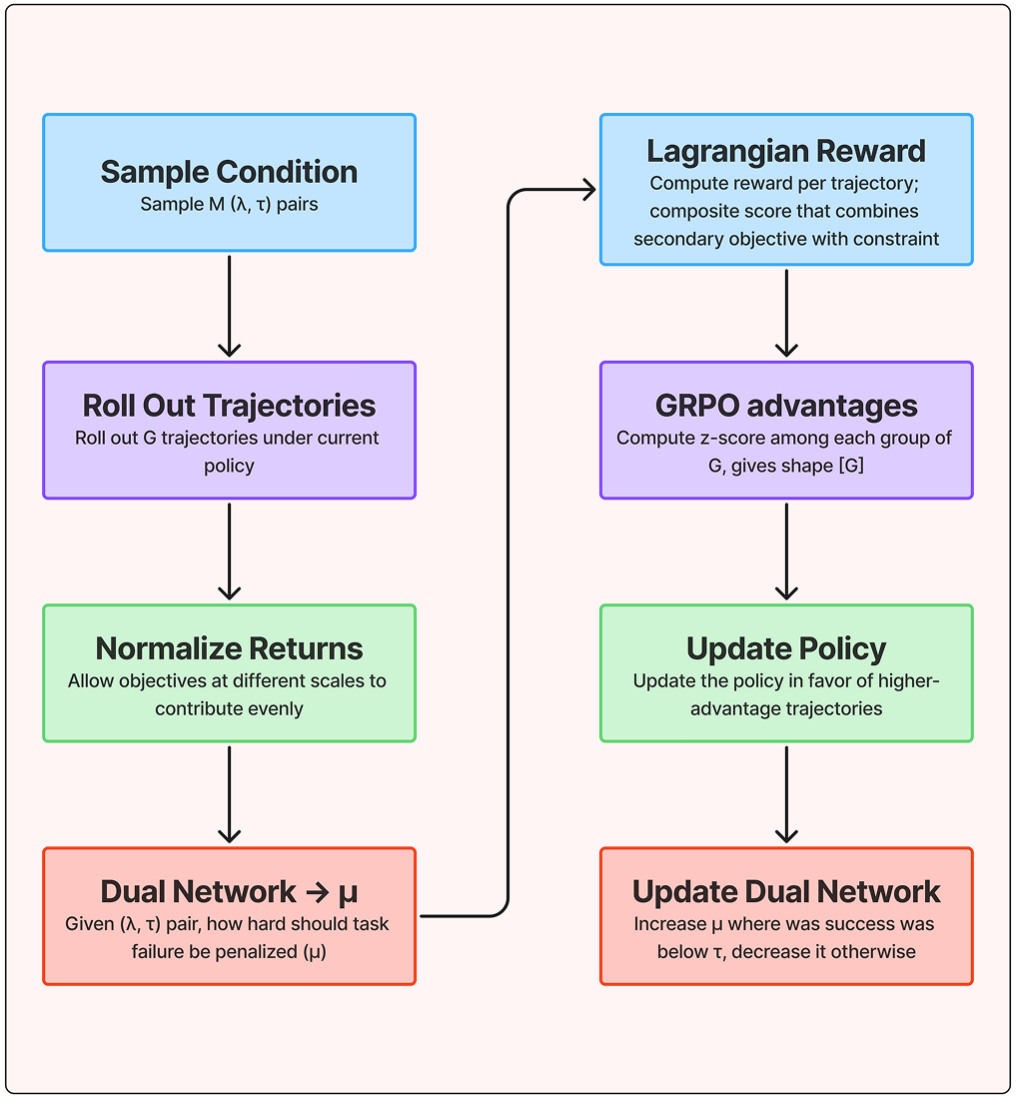
WEEK 7 UPDATE

CODY HILL • 7/6/26
DR. YUE WANG & ZAIN FARHAT



Problem Overview



Task

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

Problem Overview

	Bad	Feasible	Optimal
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

$\tau = 0.8$

Progress: SFT

- Fine-tuned base VLM
- Four task policies
- Trained on demonstrations
- High next-token prediction
- Low loss

My Results

Suite	Next-token Prediction	Steps
Spatial	95%	40,000
Object	90%	40,000
Goal	87%	40,000
Long	88%	90,000

Progress: Base SFT Evaluation

- Strong for 3/4 suites
- Issue with LIBERO Long
- Recalibrated contact forces

My Results

Suite	Avg. Success
Spatial	64%
Object	70%
Goal	54%
Long	2%

SFT Rollouts



LIBERO Object



LIBERO Spatial

Progress: CMO-GRPO Evaluation

My Results

$\lambda = (0.33, 0.33, 0.33)$

Suite	Success Floor	Success	Reward Delta
Spatial	55%	72.5% → 60%	2/3 ↓
Object	60%	57.5% → 72.5%	2/3 ↑
Goal	45%	60% → 57.5%	2/3 ↑

CMO-GRPO Rollouts



Base SFT

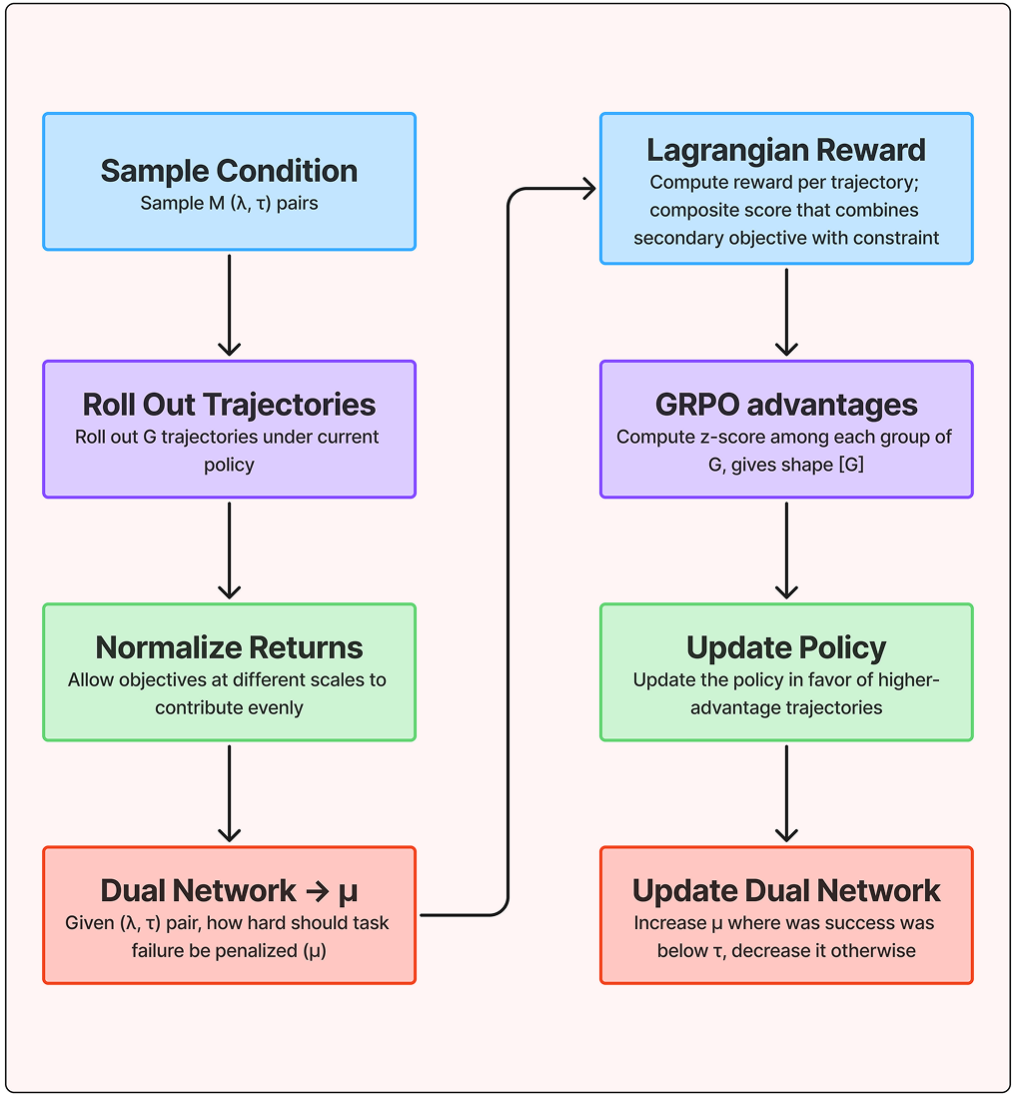


CMO-GRPO

Next Steps

Next Steps

- Currently sweeping object
- True implementation
- Clean up SFT
- Paper drafting
- Real robot



WEEK 7 UPDATE

CODY HILL • 7/6/26
DR. YUE WANG & ZAIN FARHAT

