COOPERATIVE MINGLE

KOVÁCS IMRE LEVENTE
MAKÓ KRISTÓF ZSOLT
TÓTH BENJÁMIN

MOTIVATION

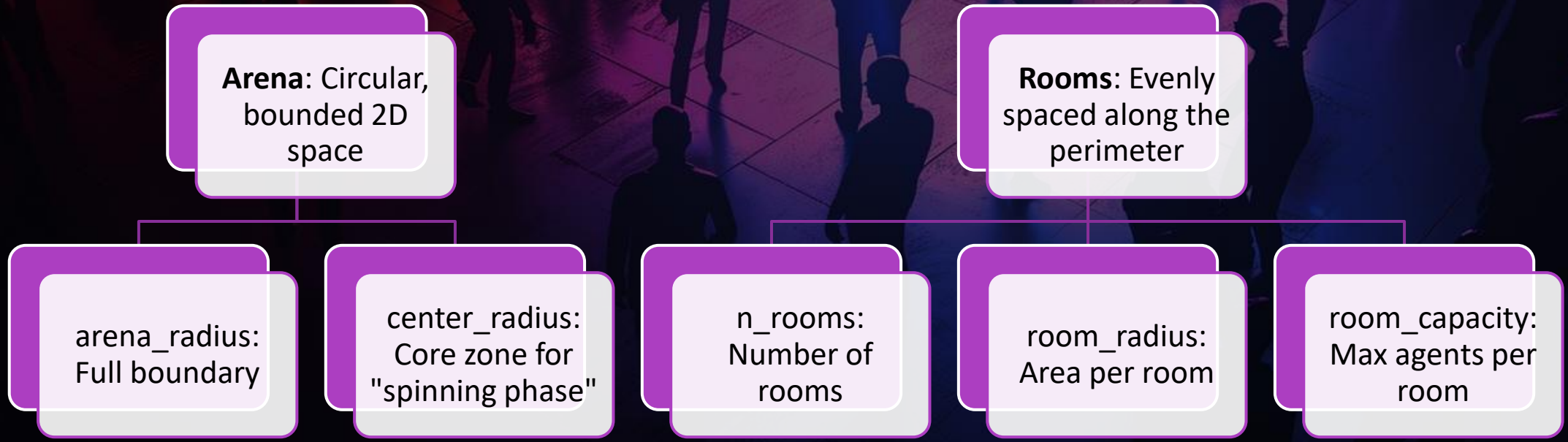
Agents **navigate dynamic phases**: spinning vs. claiming

Designed to simulate:

- Cooperative decision-making under pressure
- Spatial negotiation and coordination
- Emergent group formation strategies

Source: Netflix

ENVIRONMENT LAYOUT



PHASES

Spinning

- Agents move in the center circle
- Occasional rotation simulates social disorientation

Claiming

- Agents navigate to rooms
- Must claim spots within room capacity

AGENT OBSERVATIONS

- Distance and direction to center
- Nearest room's distance, direction, and occupancy
- Nearest agent’s distance & direction
- Phase indicator (binary flag)

REWARDS

01

All reward logic inherits from a **base RewardModule**.

02

Activated per phase: "spinning", "claiming", or "both".

03

Dynamically toggled via .active flag.

REWARDS

CloseToCenterReward

Move closer to arena center

InsideCenterReward

Stay inside center (reward/penalty)

CollisionAvoidance

Maintain safe distance from others

GetToRoomReward

Approach nearest available room

StayInRoomReward

Remain in non-overfilled room

REWARD MANAGER

Manages **progressive activation** of reward modules during training.

Enables **curriculum-style reward shaping** based on agent performance.

METRICS

CollisionRateMetric
Ratio of agent pairs too close together

RoomOccupancyRateMetric
Fraction of agents inside defined room areas

CenterPresenceMetric
Time agents spend near the environment center

AverageStepDistanceMetric
Mean distance moved per step

IdleAgentRateMetric
Fraction of agents barely moving across steps

RoomSwitchesMetric
How often agents switch closest room target

PhaseTimeMetric
Proportion of time spent in each environment phase

AgentDensityMetric
Average number of neighbors within a fixed radius

MaxDistanceFromCenterMetric
Maximum distance an agent reaches from center

MinAgentDistanceMetric
Minimum distance recorded between any two agents

AverageRoomDistanceMetric
Mean distance from agents to nearest room

AgentMovementVarianceMetric
Variance in agent movement magnitudes over time

TRAINING COMPONENTS

Custom Env via make_env()

With optional RewardManager & RewardModule

ObservationNorm

TRAINING COMPONENTS

Data Flow

Collector:
SyncDataCollector

- Gathers batches from env

Replay Buffer:
LazyTensorStorage

- Sampling with SamplerWithoutReplacement

TRAINING COMPONENTS

Learning Modules

Advantage Estimation:

StableGAE

(Generalized Advantage Estimation)

Loss:

ClipPPOLoss (with entropy bonus)

tensor(0.2243, device='cuda:0', grad_fn=<SqueezeBackward1>)

tensor(109.6843, device='cuda:0', grad_fn=<AddBackward0>)

tensor(0.2578, device='cuda:0', grad_fn=<SqueezeBackward1>)

tensor(101.8417, device='cuda:0', grad_fn=<AddBackward0>)

tensor(0.3262, device='cuda:0', grad_fn=<SqueezeBackward1>)

tensor(108.4116, device='cuda:0', grad_fn=<AddBackward0>)

tensor(inf, device='cuda:0', grad_fn=<SqueezeBackward1>)

tensor(inf, device='cuda:0', grad_fn=<AddBackward0>)

tensor(nan, device='cuda:0', grad_fn=<SqueezeBackward1>)

tensor(nan, device='cuda:0', grad_fn=<AddBackward0>)

tensor(nan, device='cuda:0', grad_fn=<SqueezeBackward1>)

tensor(nan, device='cuda:0', grad_fn=<AddBackward0>)

tensor(nan, device='cuda:0', grad_fn=<SqueezeBackward1>)

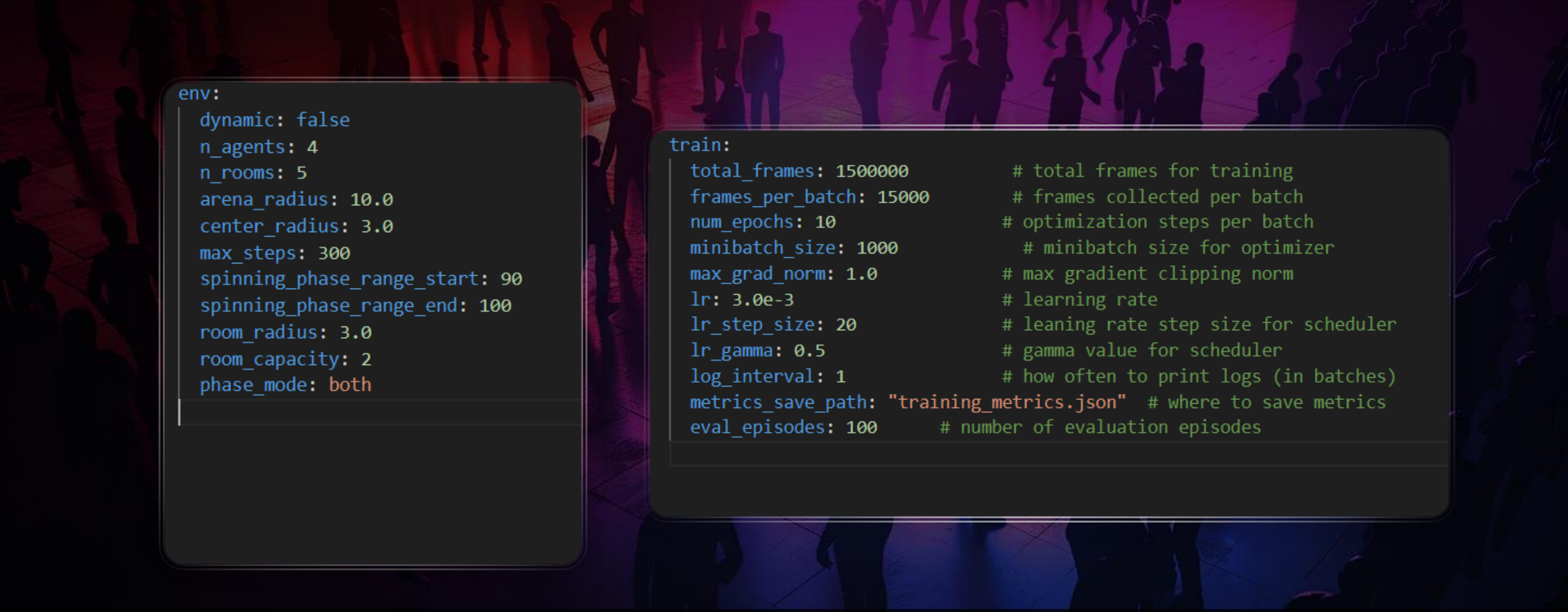
tensor(nan, device='cuda:0', grad_fn=<AddBackward0>)

tensor(nan, device='cuda:0', grad_fn=<SqueezeBackward1>)

tensor(nan, device='cuda:0', grad_fn=<AddBackward0>)

tensor(nan, device='cuda:0', grad_fn=<SqueezeBackward1>)

COMPLICATIONS


env:
dynamic: false
n_agents: 4
n_rooms: 5
arena_radius: 10.0
center_radius: 3.0
max_steps: 300
spinning_phase_range_start: 90
spinning_phase_range_end: 100
room_radius: 3.0
room_capacity: 2
phase_mode: both
train:
total_frames: 1500000 # total frames for training
frames_per_batch: 15000 # frames collected per batch
num_epochs: 10 # optimization steps per batch
minibatch_size: 1000 # minibatch size for optimizer
max_grad_norm: 1.0 # max gradient clipping norm
lr: 3.0e-3 # learning rate
lr_step_size: 20 # leaning rate step size for scheduler
lr_gamma: 0.5 # gamma value for scheduler
log_interval: 1 # how often to print logs (in batches)
metrics_save_path: "training_metrics.json" # where to save metrics
eval_episodes: 100 # number of evaluation episodes
YET ANOTHER MARKUP LANGUAGE


📦 Batch 1 - Collected 300/1200 frames
📊 Stats [Batch 1]
🕒 Time Elapsed: 39.5s
🎯 Avg Reward: -15.7667
🎭 Actor Loss: -0.0250
🏛️ Critic Loss: 0.3932
🔥 Entropy Loss: -0.0025
📉 Learning Rate: 0.003000
⚠️ Invalid Subbatches: 0/6 (0.0%)
📦 Batch 2 - Collected 600/1200 frames
📊 Stats [Batch 2]
🕒 Time Elapsed: 43.4s
🎯 Avg Reward: -9.5125
🎭 Actor Loss: -0.0116
🏛️ Critic Loss: 0.9816
🔥 Entropy Loss: -0.0025
📉 Learning Rate: 0.003000
⚠️ Invalid Subbatches: 0/6 (0.0%)
🚀 Starting training script
📂 Loading configs from folder: configs/
✅ Configs loaded and merged successfully
💻 Using device: cuda
🧱 Building training components...
Do you want to configure rewards manually? (y/n):

LOGGING

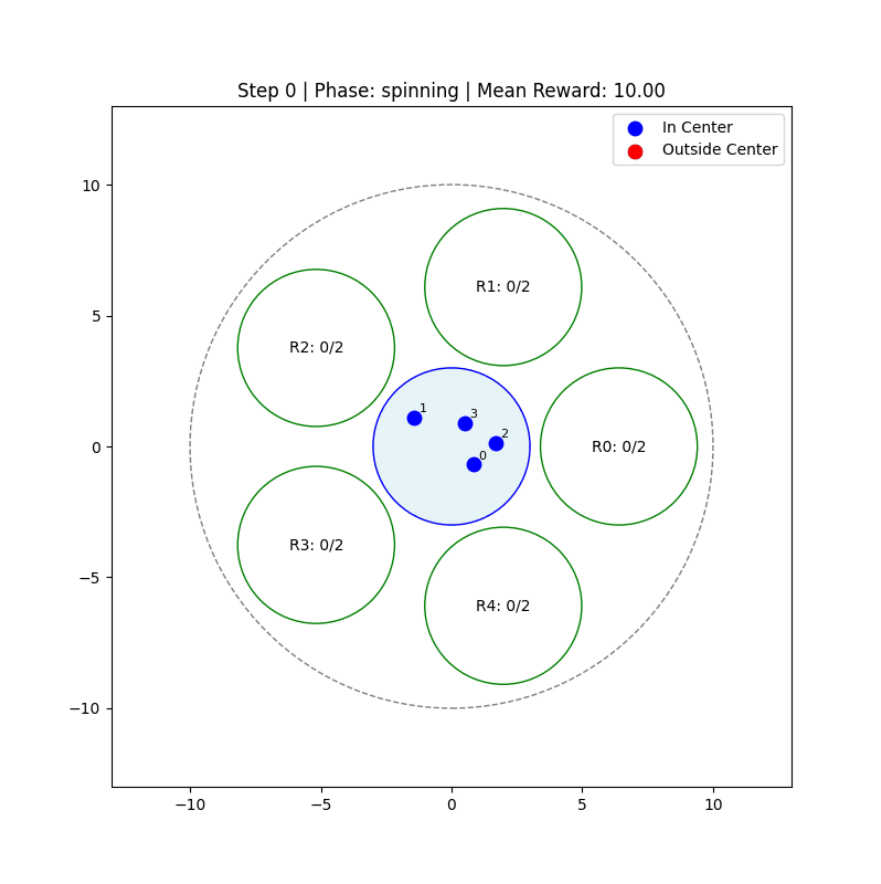
LOGGING

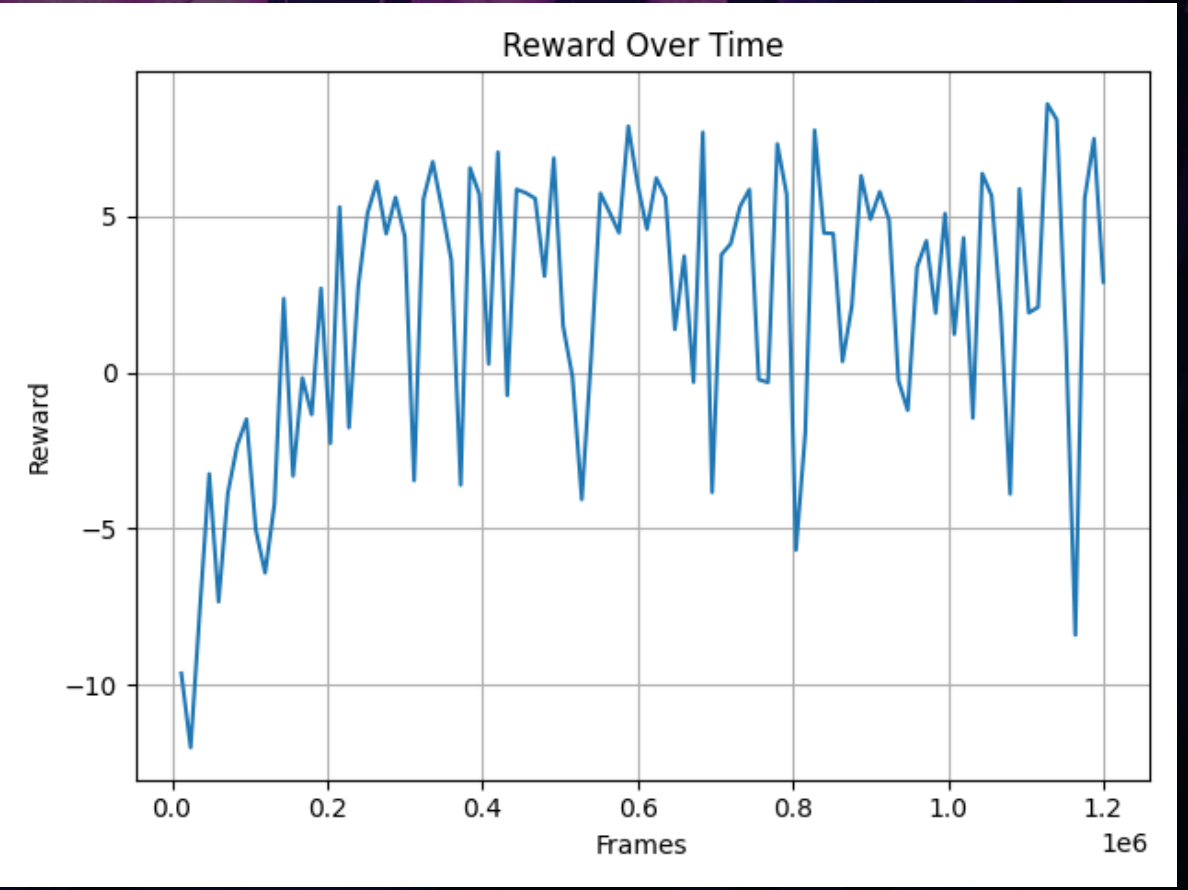
```
📈 Evaluation Summary:

🔹 collision_rate:
   Mean = 0.0134
   Std  = 0.0139

🔹 room_occupancy_rate:
   Mean = 0.2766
   Std  = 0.1315

🔹 center_presence_rate:
   Mean = 0.5590
   Std  = 0.1522

🔹 average_step_distance:
   Mean = 0.2056
   Std  = 0.0023
```

```
🔹 average_room_distance:
   Mean = 3.7722
   Std  = 0.4310

🔹 agent_movement_variance:
   Mean = 0.0023
   Std  = 0.0006

✅ Training & Evaluation finished successfully

💾 Saved training metrics to training_metrics.json
```

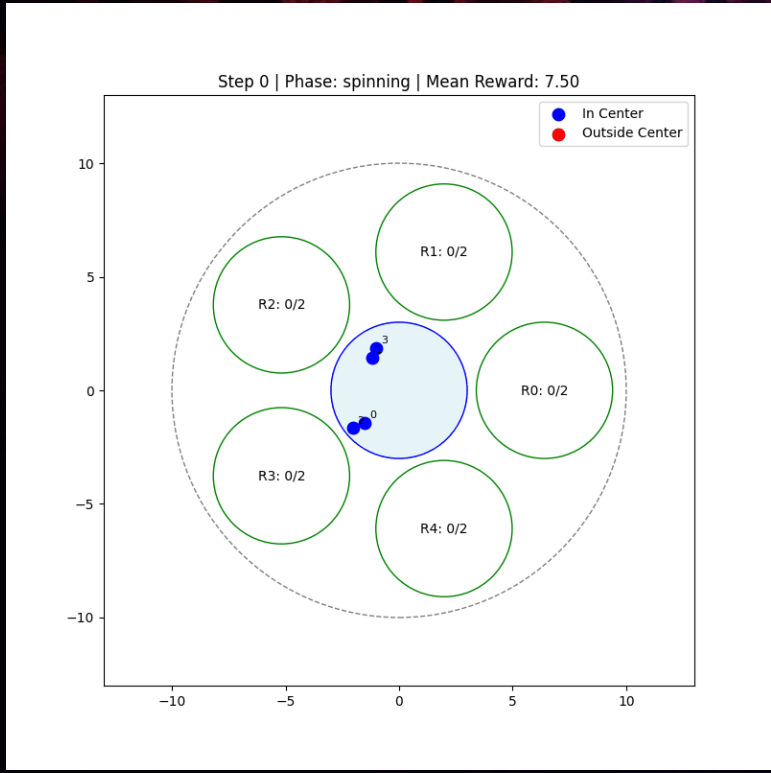
```
🎯 Episode 100/100 results:
   collision_rate: 0.024
   room_occupancy_rate: 0.119
   center_presence_rate: 0.748
   average_step_distance: 0.207
   idle_agent_rate: 0.000
   room_switches: 53.000
   phase_time_spinning: 0.320
   phase_time_claiming: 0.680
   agent_density: 0.187
   max_distance_from_center: 7.244
   min_agent_distance: 0.091
   average_room_distance: 4.340
   agent_movement_variance: 0.002

📁 Metrics saved to eval_results\eval_metrics_20250601_180322/metrics.json

📊 Plots saved to: eval_results\eval_metrics_20250601_180322
```

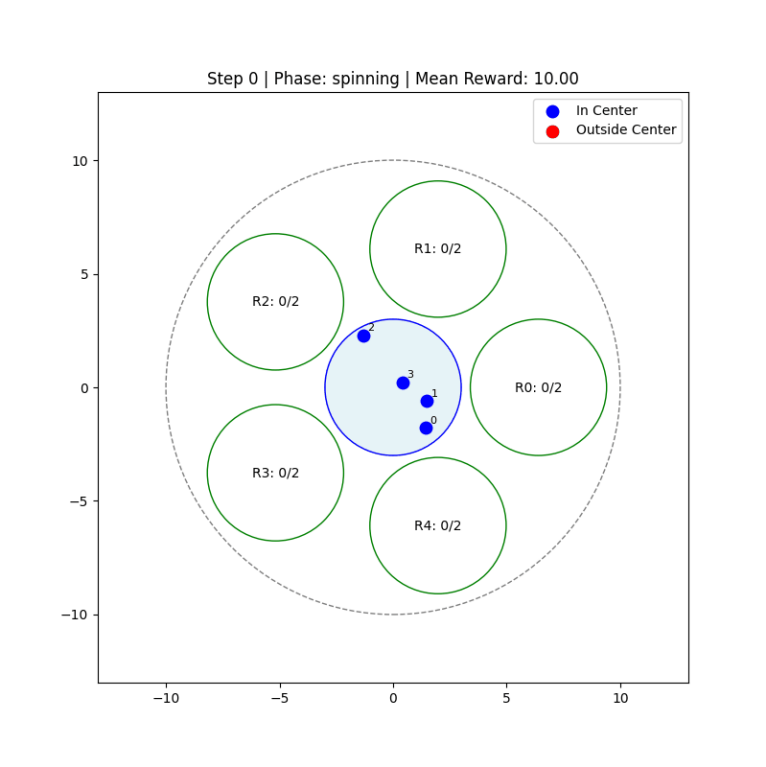
RESULTS

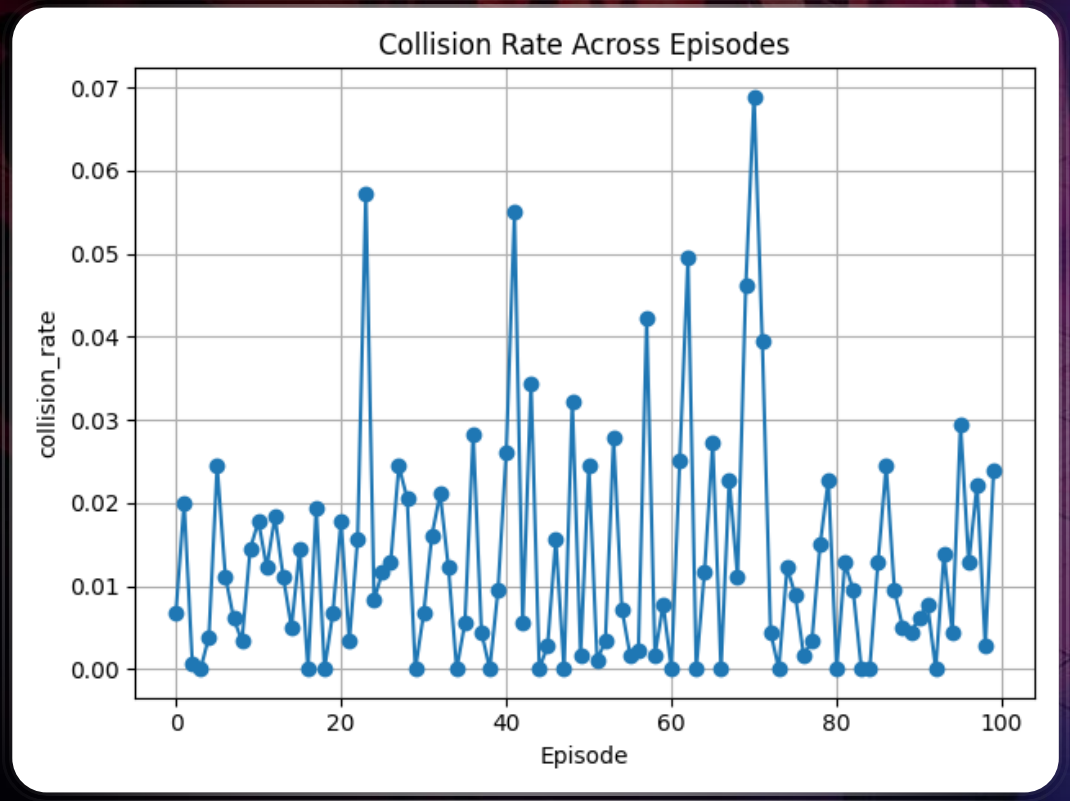

Step 0 | Phase: spinning | Mean Reward: 10.00
In Center
Outside Center
R1: 0/2
R2: 0/2
R0: 0/2
R3: 0/2
R4: 0/2
10
5
0
−5
−10
−10
−5
0
5
10

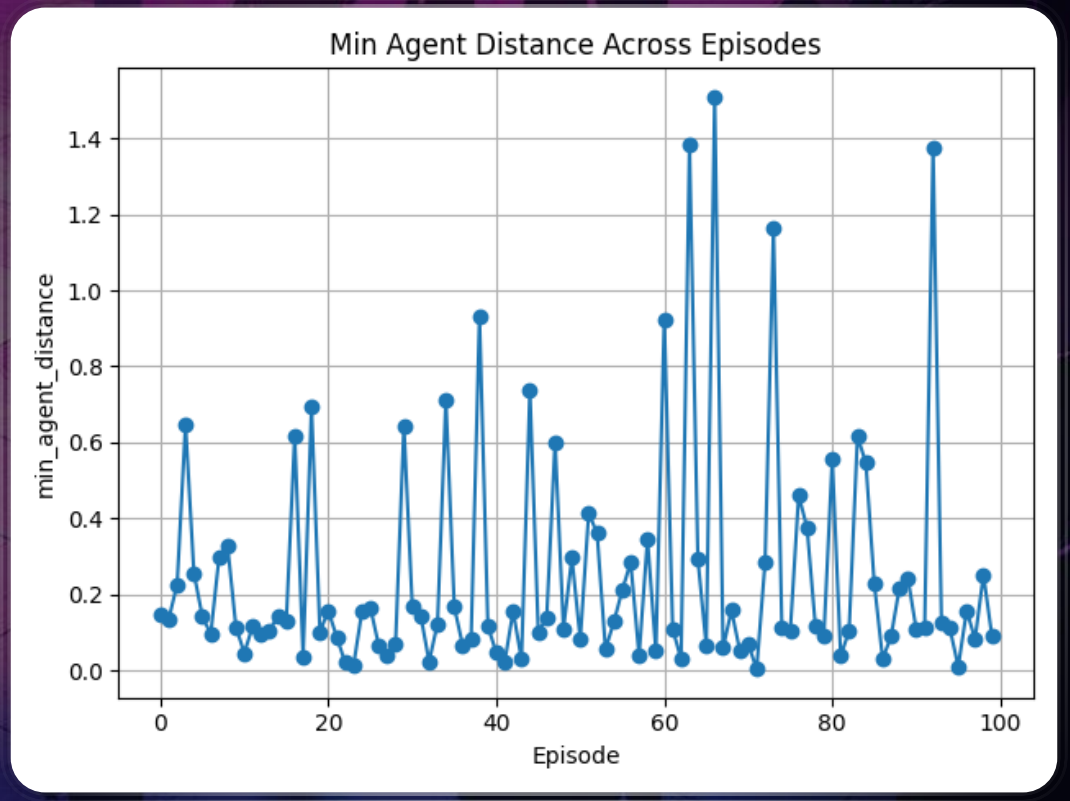

Reward Over Time
Reward
5
0
−5
−10
0.0
0.2
0.4
0.6
0.8
1.0
1.2
Frames
1e6

RESULTS


Step 0 | Phase: spinning | Mean Reward: 7.50
In Center
Outside Center
R1: 0/2
R2: 0/2
R0: 0/2
R3: 0/2
R4: 0/2
10
5
0
−5
−10
−10
−5
0
5
10

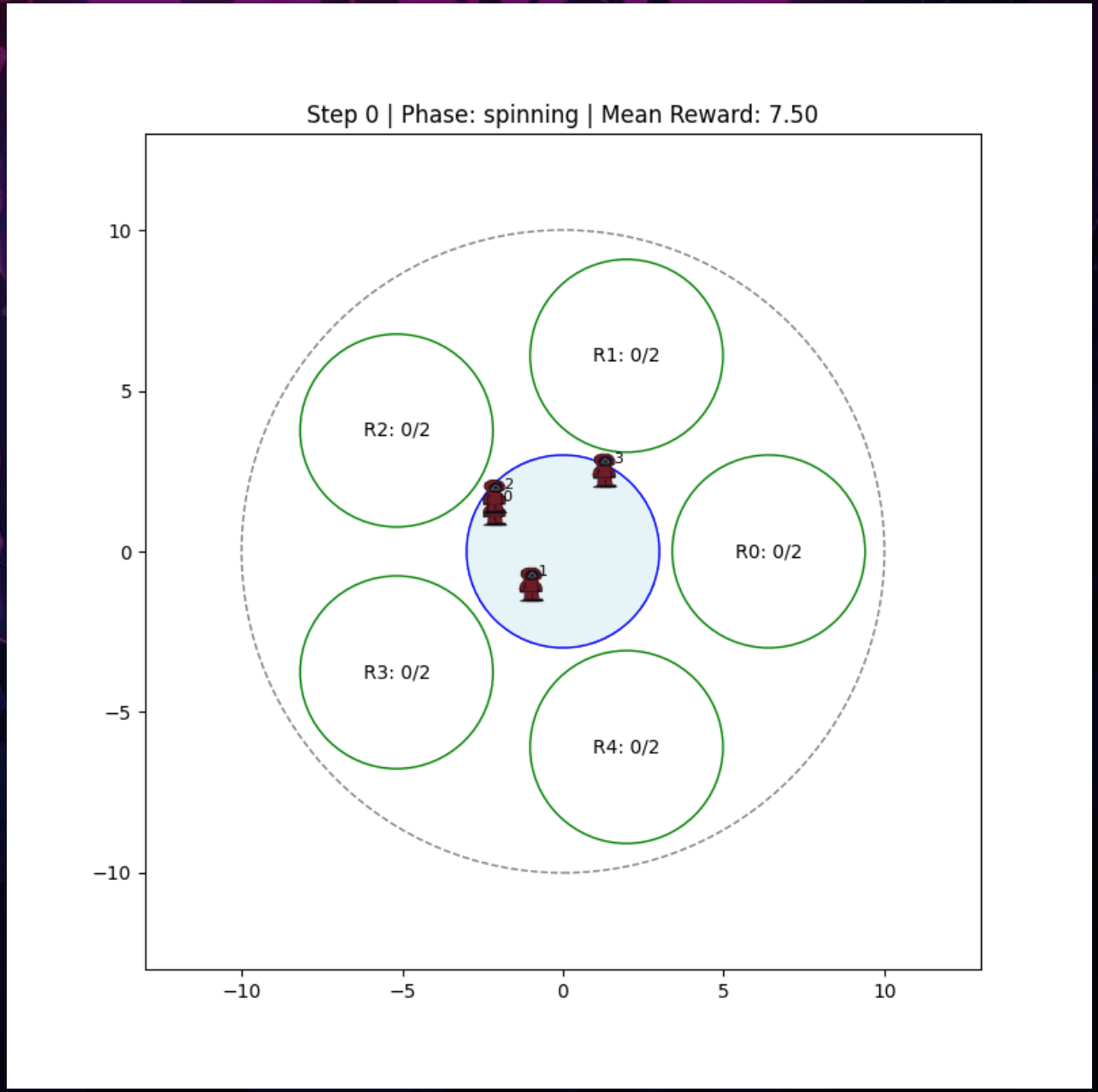

Step 0 | Phase: spinning | Mean Reward: 10.00
In Center
Outside Center
R1: 0/2
R2: 0/2
R0: 0/2
R3: 0/2
R4: 0/2
10
5
0
−5
−10
−10
−5
0
5
10


Collision Rate Across Episodes
collision_rate
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0
20
40
60
80
100
Episode


Min Agent Distance Across Episodes
min_agent_distance
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
20
40
60
80
100
Episode

RESULTS

FUTURE DIRECTIONS



THANK YOU FOR YOUR ATTENTION!