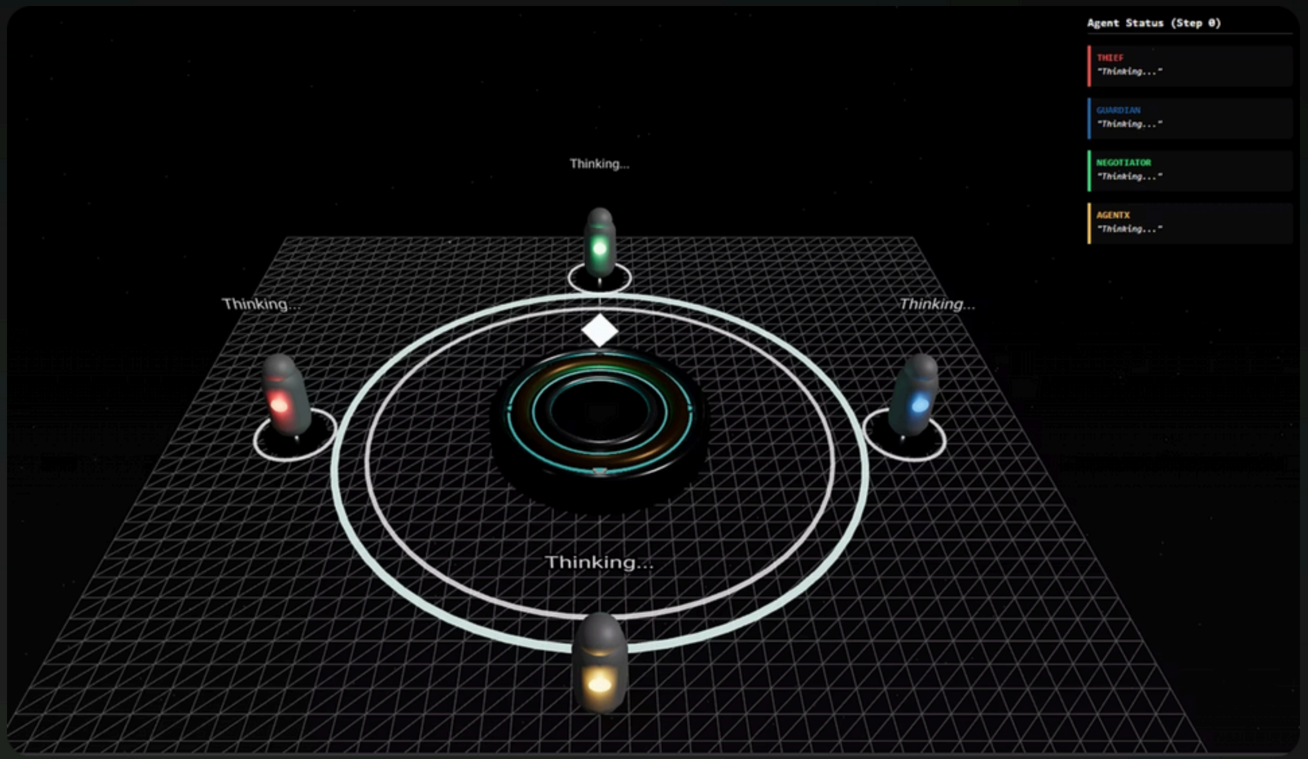

Multi-Agent Reinforcement Learning

Crystal Heist Simulation

Team NGen

V. Jaya Sai Reddy • Songa Kiranmai • K. Srishanth Reddy • K. Hemanth

🌐 **Project Website:** crystal-hackathon.s3-website-us-east-1.amazonaws.com

Problem Statement

Investigating three critical aspects of Multi-Agent Reinforcement Learning systems through experimental simulation:

Interaction Quality

Analyzing how agents interact in shared environments, examining both cooperative and competitive behavioral patterns and strategic decision-making processes.

Emergent Behavior

Identifying and documenting emergent behaviors arising from agent interactions, including unexpected strategies, collective patterns, and adaptive responses.

Reward Fairness

Evaluating fairness of reward distribution among multiple agents to ensure balanced learning outcomes, equitable performance, and sustainable cooperation.

System Flow & Function

Multi-Agent Environment: Crystal Heist

Agent Roles

- 🥷 Thief: Attempts to steal the crystal, must coordinate with others to open vault through strategic plate activation
- 🛡️ Guardian: Protects the vault while adapting between cooperative and competitive strategies based on situation
- 🤝 Negotiator: Facilitates cooperation through strategic positioning and implicit communication signals
- ❓ AgentX: Wild card agent that learns adaptive strategies, exploring novel behavioral patterns

Learning Pipeline

1. **Initialize**
   Set Q-values, parameters
2. **Interact**
   Observe, Decide, Act
3. **Learn**
   Update Q-values
4. **Evaluate**
   Measure performance

Action Space: go_plate_0/1/2 (requires 3 agents), go_vault, wait

Algorithm: Q-Learning

Q-Learning Update Formula

$$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$$

Hyperparameters

- **α (alpha) = 0.2** → Learning rate controls update magnitude
- **γ (gamma) = 0.95** → Discount factor for future rewards
- **ε (epsilon) = 0.2** → Exploration rate balancing exploration-exploitation

Epsilon-Greedy Strategy

- **20% exploration** → Random action selection for discovering new strategies
- **80% exploitation** → Leveraging best known actions from learned Q-values

Reward Structure

- **+0.2** when vault opens (cooperation incentive)
- **+50.0** for crystal obtained (goal achievement)
- **-0.05** per step (efficiency penalty to minimize episode length)

Results & Product Demo

Training Performance Metrics

10+

Episodes

Training iterations

75%

Success Rate

Vault opening frequency

300%

Improvement

vs Random Baseline

🎮 **Live Demo Features:** Real-time 3D visualization, training metrics dashboard, interactive agent controls, Q-value heatmaps

Learned Policy vs Random Baseline

Learned: High success (>70%), efficient coordination, reduced steps to goal, strategic positioning

Baseline: Low success (<25%), inefficient coordination, high variance, poor strategic planning

Multi-Agent Interaction & Emergent Behavior

Cooperation Evidence

- **Coordinated plate activation:** 3 agents simultaneously positioning on plates
- **Role specialization:** Agents develop preferences for specific plates and positions
- **Temporal synchronization:** Agents learn optimal timing for joint actions

Competition & Strategy

- **Race dynamics:** Competition to reach vault first for +50 reward
- **Nash Equilibrium:** Cooperate to open vault, then compete for crystal acquisition
- **Strategic trade-offs:** Balancing cooperation cost vs competitive advantage

Observable Emergent Behaviors

- Strategic positioning near vault while maintaining plate activation
- Implicit communication through learned positional patterns
- Adaptive waiting strategies for improved coordination timing
- Dynamic role switching based on environmental state

Fairness Analysis

✓ **Shared rewards** (+0.2 for all agents when vault opens) • ✓ **Individual rewards** (+50 for crystal obtainer) • ✓ **Balanced learning** across all agent policies • ✓ **Equal step penalties** (-0.05) ensure efficiency focus

Key Achievements & Insights

✅ Successfully Demonstrated

- **Multi-agent coordination in cooperative-competitive environments with complex strategic trade-offs**
- **Q-learning with epsilon-greedy exploration achieving convergence to near-optimal policies**
- **Emergent strategic behaviors arising from simple reward structures without explicit programming**
- **Fairness in distributed rewards ensuring equitable learning opportunities for all agents**
- **Real-time 3D visualization enabling intuitive understanding of agent learning dynamics**

Key Insights

Cooperation emerges naturally when reward structures properly incentivize collaborative behavior. Competition drives efficiency optimization and strategic refinement. The balance between cooperation and competition creates rich, adaptive multi-agent behaviors that exceed individual agent capabilities.

Tech Stack: React + Three.js | TypeScript | Q-Learning RL | AWS S3

Thank You!

Team NGen