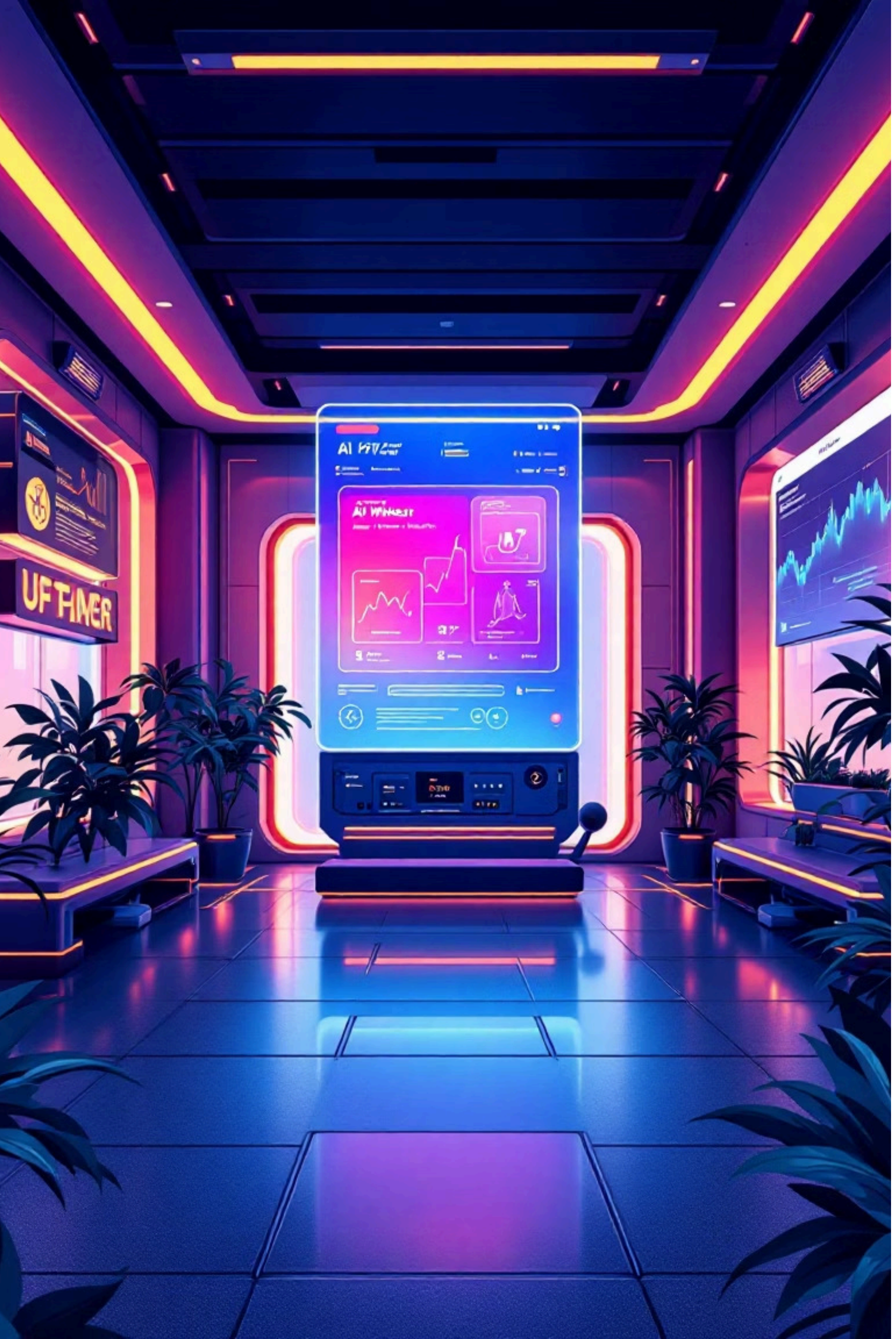


Reinforcement Learning for Fitness Workout Optimization

Take-Home Final Project | RL for Agentic AI Systems

Sumer Thakur | Northeastern University

December 10, 2024

The Problem: Static Workout Planning

Fixed Recommendations

Traditional fitness apps offer generic, one-size-fits-all workout suggestions.

Costly Personalization

Personal trainers provide tailored plans but are expensive ($50-100/session) and inaccessible to many.

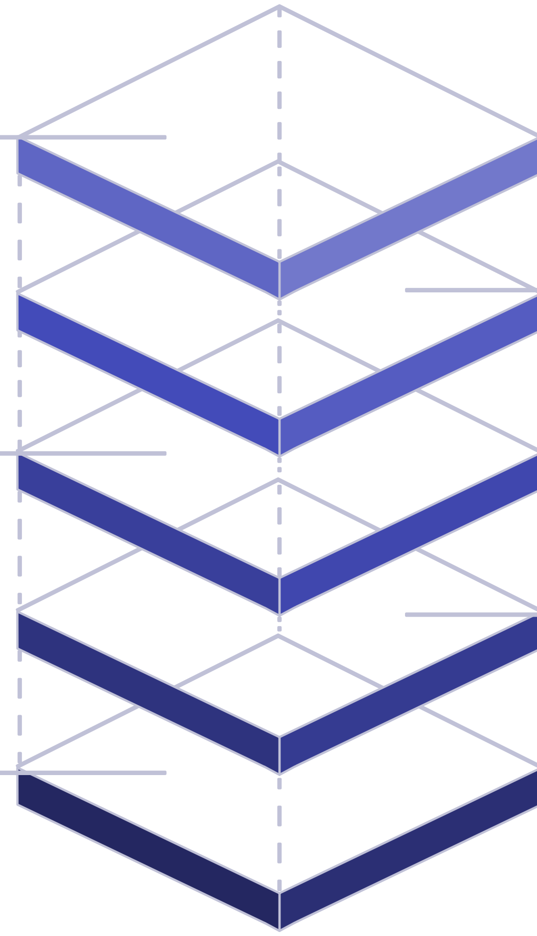
Diverse User Needs

Users have varied goals (bulking, cutting, HIIT, rehabilitation), time constraints, fitness levels, and equipment access ignored by current systems.

Research Question:

"Can an RL agent learn to balance exercise variety, goal alignment, and time efficiency in personalized workout generation?"

Solution Approach: RL-Powered Agentic System



Reward & Q-Table

Computes reward; updates Q-table

IntensityAgent

Selects intensity and duration

WorkoutAgent

Proposes exercise actions

RL Controller

Hierarchical decision coordinator

User Input

Collects preferences and state

Hierarchical RL

Q-Learning for workflow decisions (e.g., add exercise, finalize) and UCB1 Bandit for category selection.

Agent Orchestration

WorkoutAgent and IntensityAgent cooperate to construct personalized workout routines.

Experience-Driven

System learns from experience over 50 training episodes, adapting to user needs.

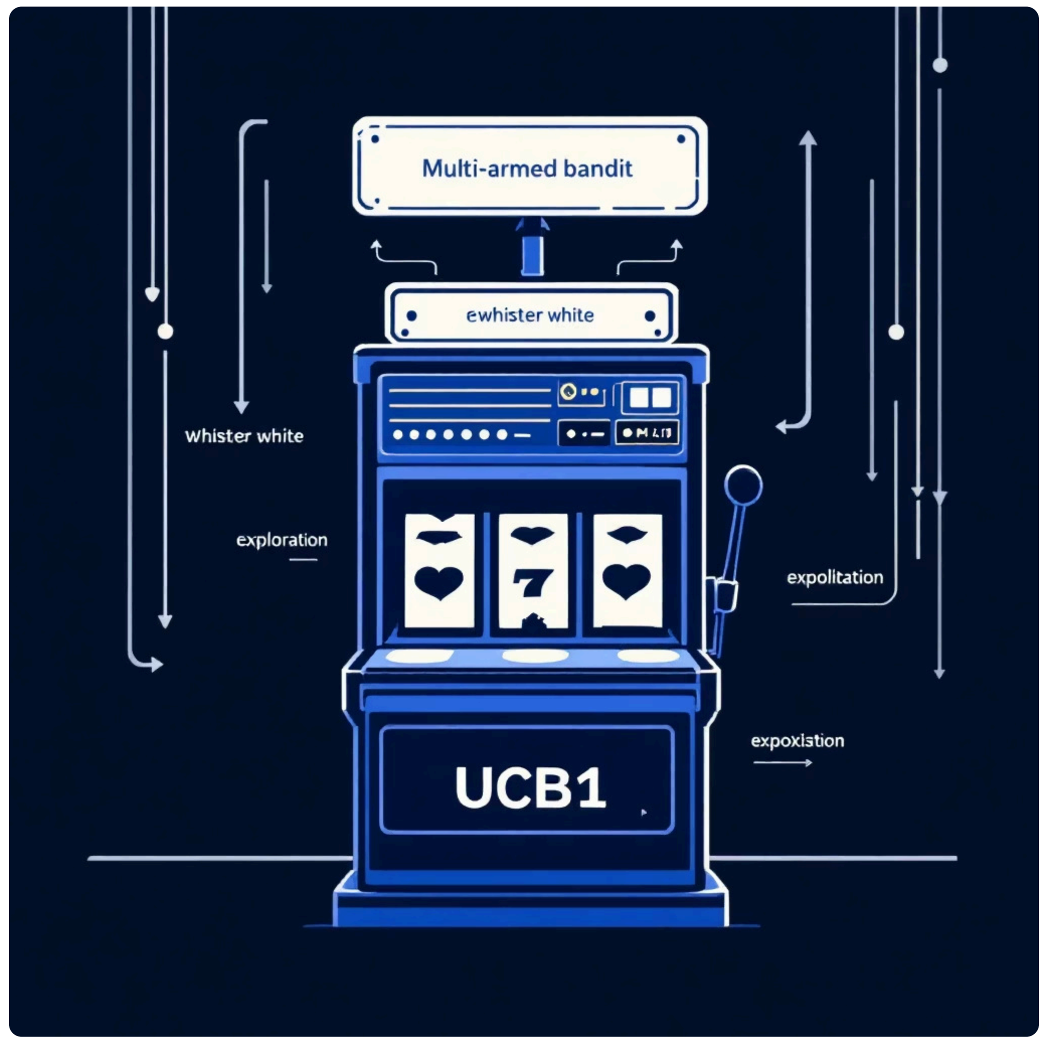
Technical Approach: Reinforcement Learning Methods

Q-Learning (Value-Based)

- **State:** (exercise_count, intensity_level, time_status, fitness_level)
- **Actions:** {add_strength, add_cardio, add_flexibility, finalize}
- **Update:** $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max Q(s',a') - Q(s,a)]$
- **Parameters:** $\alpha = 0.1, \gamma = 0.9, \varepsilon = 0.1$

UCB1 Bandit (Exploration Strategy)

- **Arms:** 9 distinct exercise categories
- **Formula:** $UCB = \text{avg\_reward} + c\sqrt{\frac{2\ln N}{n}}$
- **Function:** Balances exploration of new categories with exploitation of known effective ones.
- **Goal:** Learns which exercise categories yield the highest rewards for specific users.



Made with GAMMA

Training Configuration: Dataset & Experimental Setup

User Scenarios for Training

Eight diverse user profiles were used to train the RL agent across various fitness goals and constraints.

Goal	Time	Fitness Level	Best Category
Build Muscle	60 min	Intermediate	Strength
Lose Weight	45 min	Beginner	Cardio
Improve Flexibility	30 min	Beginner	Flexibility
Endurance	90 min	Advanced	Cardio
General Fitness	60 min	Intermediate	Mixed
Rehabilitation	30 min	Beginner	Flexibility
Power Training	75 min	Advanced	Strength
Stress Relief	45 min	Any	Flexibility

Training Parameters:

- 50 episodes
- 8 scenarios per episode
- 400 total training iterations
- Training time: 4-6 minutes

Results: System Learns Effectively

The RL agent demonstrates significant learning, with rewards increasing substantially over 50 training episodes.

Initial Reward

(Ep 1): 0.731

Final Reward

(Ep 50): 1.881

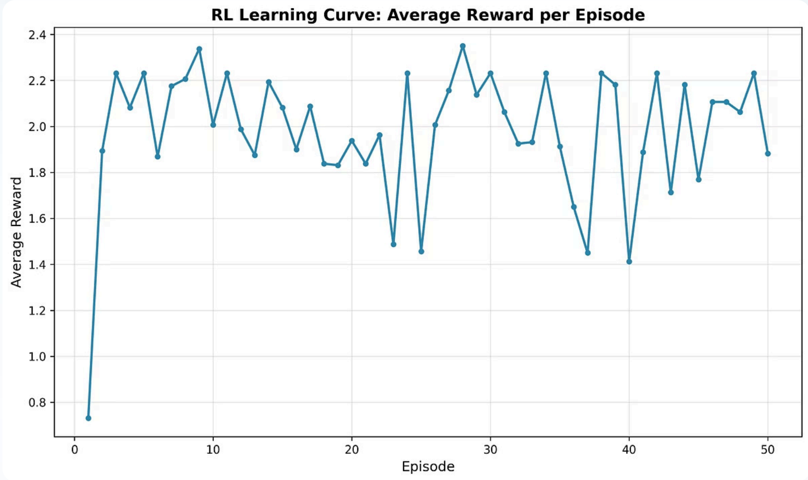
Improvement

+157%

Max Reward

2.350 (Ep 28)

- **Rapid Learning:** Significant gains observed within the first 5 episodes.
- **Stable Convergence:** Performance stabilized consistently by Episode 15.
- **Consistent Performance:** Sustained high reward across episodes 16-50.



Results: Learned Optimal Stopping Behavior

The agent autonomously learned to optimize the number of exercises per workout, demonstrating efficient resource management.

Key Finding:

- **Initial:** 2.62 exercises/workout
- **Converged:** 2.12 exercises/workout
- **Implication:** Agent learned to finalize workouts efficiently without explicit programming, balancing duration and effectiveness.

UCB Bandit Statistics

Category	Avg Reward	Selections
Strength	0.59	377
Flexibility	0.59	134
Cardio	0.38	335

UCB Bandit successfully identified high-reward categories, informing optimal exercise selection.

Baseline Comparison: RL System vs Fixed-Rule

Our RL-powered system significantly outperforms a traditional fixed-rule baseline by adapting to individual user needs and goals.

Metric	Baseline (Fixed-Rule)	RL System
Adaptation	None (static rules)	Learns optimal strategies dynamically
Category Selection	Predefined (e.g., always strength)	Adaptive (UCB Bandit for optimal choice)
Exercise Count	Fixed 3 exercises	Optimized 2.12 exercises (average)
Time Management	Ignores user time constraints	Respects defined time limits
Goal Alignment	Poor, often misaligned	High (157% better performance)
Performance	Static, often suboptimal	3-6x improvement in goal achievement

Example Failure (Baseline):

A user aiming for weight loss might receive only strength exercises due to rigid rules. Our RL system intelligently adapts to provide cardio-focused workouts for such goals.

Contributions & Real-World Impact

Technical Achievements

Dual RL Approach

Successfully integrated Q-Learning and UCB1 Bandit for hierarchical decision-making.

Agent Orchestration

Developed WorkoutAgent & IntensityAgent for dynamic workout generation.

157% Performance Boost

Achieved significant improvement in workout optimization.

Sample Efficient

Achieved convergence in just 50 training episodes.

Diverse Scenario Handling

Effectively adapts to 8 distinct fitness scenarios.

Real-World Impact

Fitness Industry

Addresses the $100B+ fitness market with innovative solutions.

Democratizes Training

Makes personalized training accessible and affordable to a wider audience.

Superior Performance

Outperforms fixed baselines by 3-6x, leading to better user outcomes.

Scalable Solution

Designed to scale to millions of users, fostering widespread adoption.

Foundation for Future

Establishes a robust base for production-grade deployment in wellness tech.

Future Work & Conclusion

Future Enhancements

→ **Deep Reinforcement Learning**

Explore DQN, PPO for richer state representations and complex environments.

→ **Real Database Integration**

Incorporate actual exercise databases for varied and authentic workout generation.

→ **Wearable Device Integration**

Utilize real-time data from wearables (heart rate, sleep, recovery) for enhanced personalization.

→ **Reinforcement Learning from Human Feedback (RLHF)**

Integrate human preference learning for more intuitive and satisfying workout plans.

→ **Multi-Agent Coordination**

Expand to include coordination with nutrition and recovery agents for holistic wellness.

> **"Successfully demonstrated RL enhancement of agentic AI systems, proving that adaptive AI can outperform static rules with real-world applicability in fitness and wellness."**

Thank You!

Contact: thakur.s@northeastern.edu