Movie Recommendation system

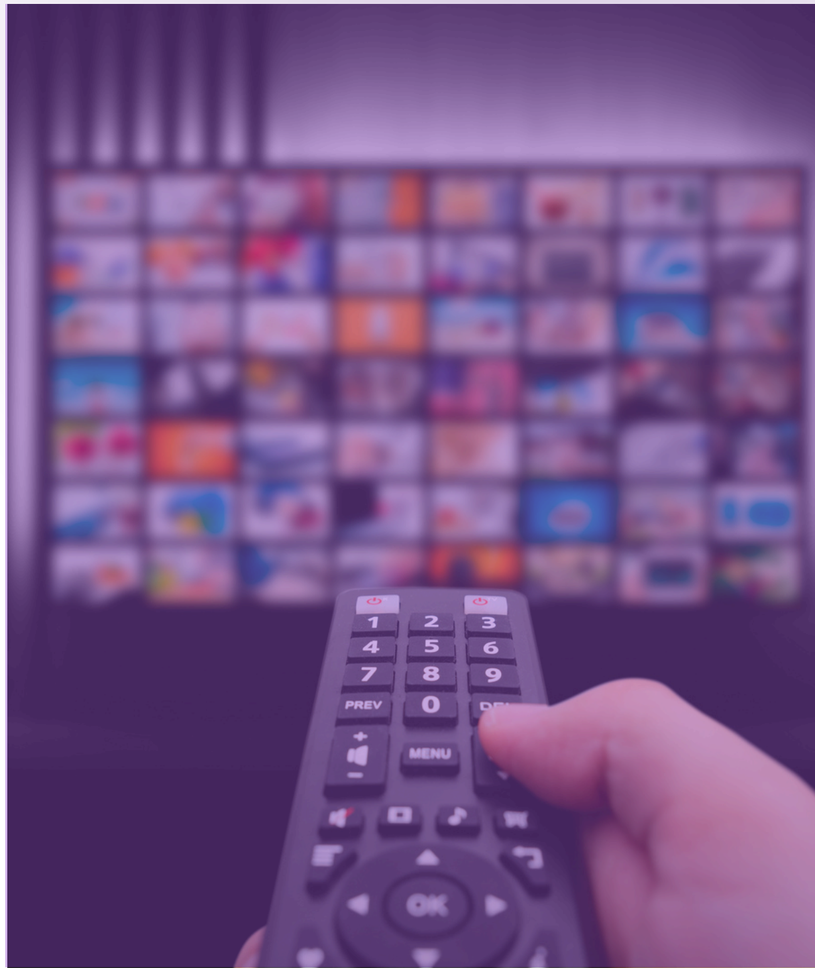
Reinforcement Learning Course

Dr. Teddy Lazebnik

Odeya Hazani
Roni Epstein

Introduction

- **Goal**: Develop an agent that **learns to recommend movies** to simulated users with changing preferences.
- **Motivation**: Personalization in recommendations is a core challenge – user interests evolve over time and depend on interaction history.
- **Approach:**
  - RL environment simulates user behavior, rewards, and dynamics.
  - **Dueling DQN** agent.
- **Deliverables:**
  - Modular Python codebase (env, agent, training, evaluation)



ENVIRONMENT DESIGN

User Profile

Gender

Age Group

Favorite Genres [%]

Boredom

Forbidden Genres

Problem Formulation

01 State

- User demographics.
- Interaction history (recent movies, ratings)
- Current slate (candidate movies)

02 Actions

Select a slate of N movies from available candidates

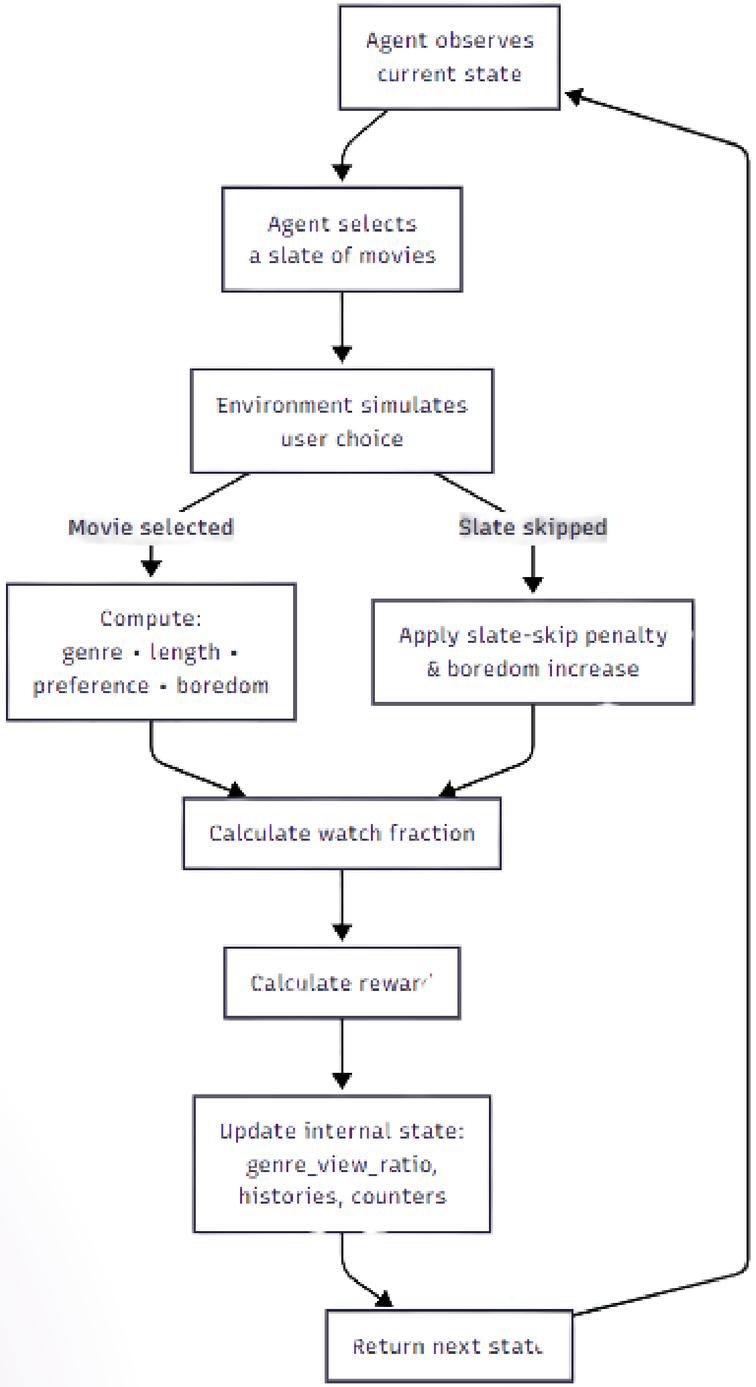
03 Reward

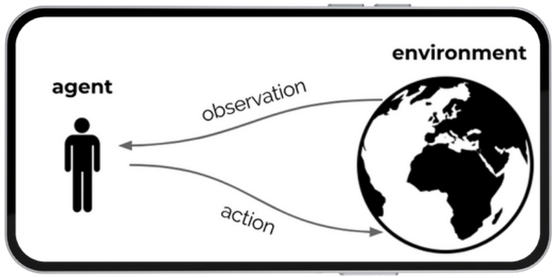
- Reward = user engagement (watch fraction)
- Positive for watched/liked movies, zero/negative otherwise

04 Environment

User preferences evolve after each interaction (stochastic update)

The interaction loop





Algorithm Overview

Baselines

Random
Selects a slate of movies randomly at each step, without considering user history or preferences.

Static
Repeats the same fixed slate of movies throughout the episode, ignoring any feedback or change in user state.

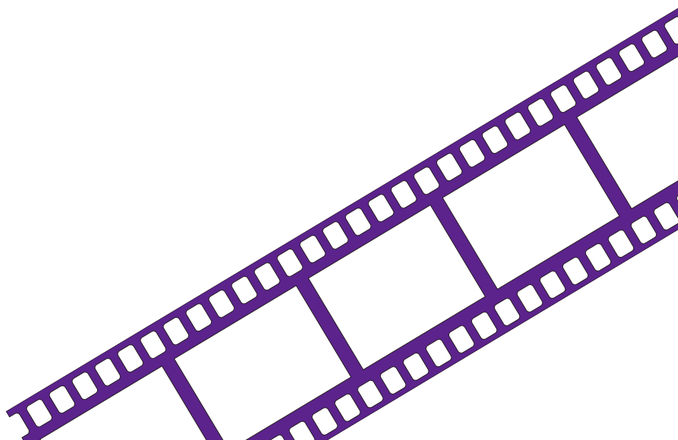
Popularity
Always recommends the most popular movies from the catalog, regardless of the specific user.

Greedy
Chooses movies with the highest immediate predicted engagement, focusing only on short-term gain without long-term planning.

Main Algorithm

Dueling DQN
(Dueling Deep Q-Network)

A value-based deep reinforcement learning algorithm that splits the estimation of the state value and the advantage of each action.



Dueling DQN

- **Scalability**: Handles complex, high-dimensional state and action spaces.
- **Generalization**: Learns across diverse users and sessions.
- **Sample Efficiency:** Focuses learning on important transitions (PER).
- **Stability:** Dueling network and target updates reduce instability.
- **Suitability:** Designed for sequential, combinatorial recommendation tasks.

EXPERIENCE REPLAY & PRIORITIZATION

Replay Buffer:

- Stores recent transitions (state, action, reward, next state) up to fixed capacity.
- Allows mini-batch sampling, breaking temporal correlations.

Prioritized Experience Replay (PER):

- Transitions are sampled with probability proportional to TD-error.
- Focuses learning on "surprising" or high-error experiences.

Importance Sampling:

- Weights applied to correct bias from prioritized sampling.

Benefits:

- Improves sample efficiency and stabilizes training in non-stationary, stochastic environments.

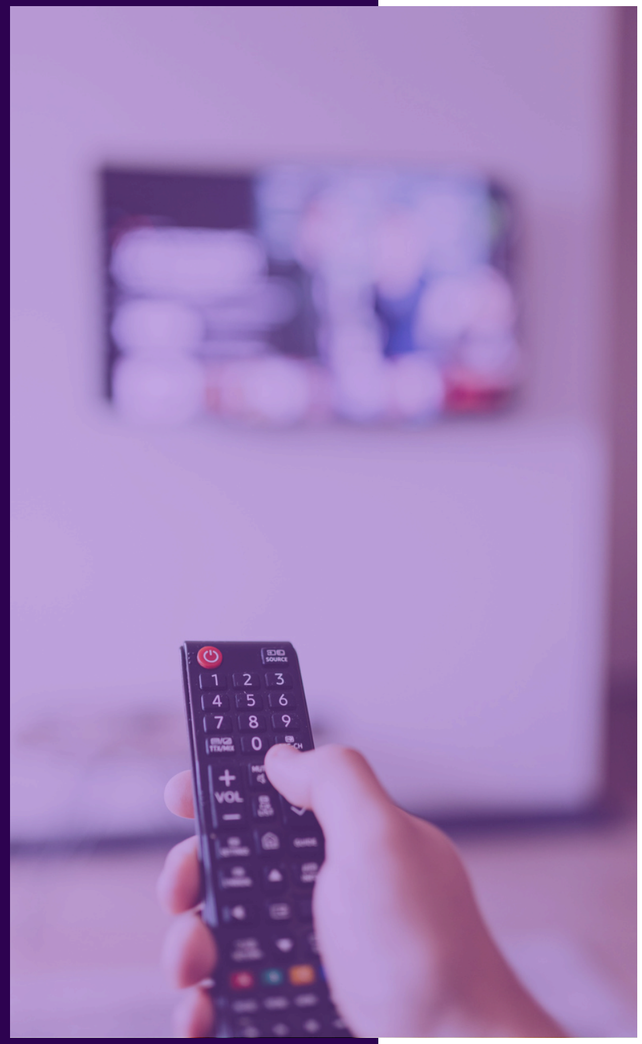
ARCHITECTURE & HYPERPARAMETERS

Neural Network Architecture

- Input:
  - Concatenated feature vector: user demographics, preference vector, interaction history, current candidate slate.
- Hidden Layers:
  - Layer 1: 256 neurons (ReLU)
  - Layer 2: 128 neurons (ReLU)
- Dueling Heads:
  - Separate value and advantage streams
- Output Layer:
  - Q-values for all possible valid slates (action masking applied)

Training Hyperparameters

- Discount Factor (γ): 0.99
- Learning Rate (LR): 0.0001
- Batch Size: 64
- Replay Buffer Size: 50,000
- Prioritized Experience Replay: Enabled (α=0.6, β annealed to 1.0)
- Exploration (ε-greedy):
  - Start: 1.0, End: 0.05, Decay: 0.995
- Target Network Update: Every 10 episodes
- Max Episode Length: 20 recommendations



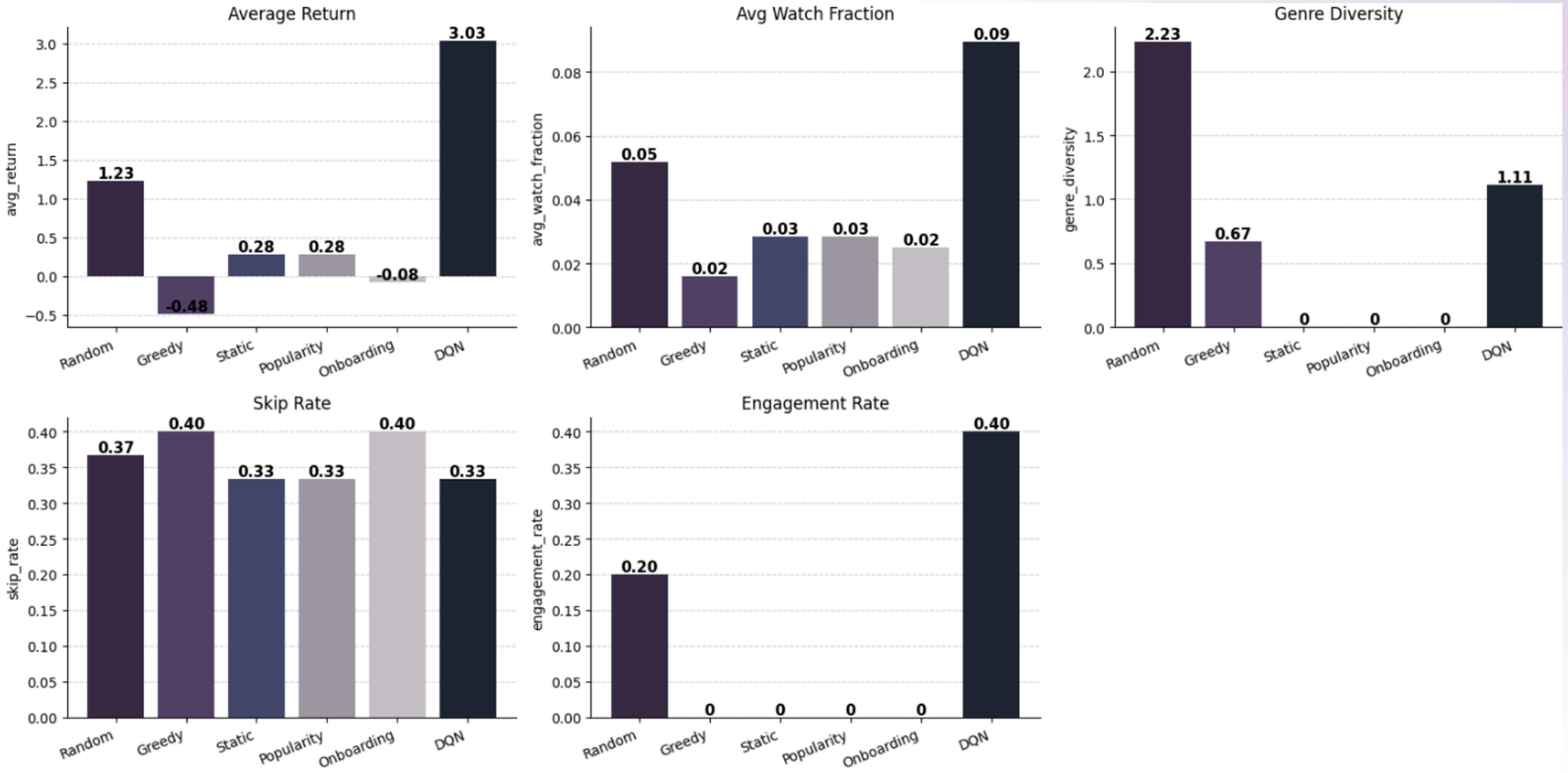
EVALUATION

Metrics

- **Average Return**: Mean cumulative reward per episode.
- **Average Watch Fraction**: Fraction of slate actually watched by user.
- **Diversity**: Number of unique genres recommended.
- **Engagement Rate:** How often users are interested enough to watch a movie beyond a minimal fraction.
- **Skip Rate:** Fraction of slates where the user skipped all recommendations, out of all presented slates.


SPOTLIGHT
THE REVENANT
DEADPOOL
THE BOY
33 HOURS
ERSCHÜTTERNDE WAHRHEIT
DIRTY GRANDPA
COLONIA DIGNIDAD
ANNE FRANK

Results


Average Return
avg_return
1.23
-0.48
0.28
0.28
-0.08
3.03
Avg Watch Fraction
avg_watch_fraction
0.05
0.02
0.03
0.03
0.02
0.09
Genre Diversity
genre_diversity
2.23
0.67
0
0
0
1.11
Skip Rate
skip_rate
0.37
0.40
0.33
0.33
0.40
0.33
Engagement Rate
engagement_rate
0.20
0
0
0
0
0.40
Random
Greedy
Static
Popularity
Onboarding
DQN

DISCUSSION & INSIGHTS

Challenges:

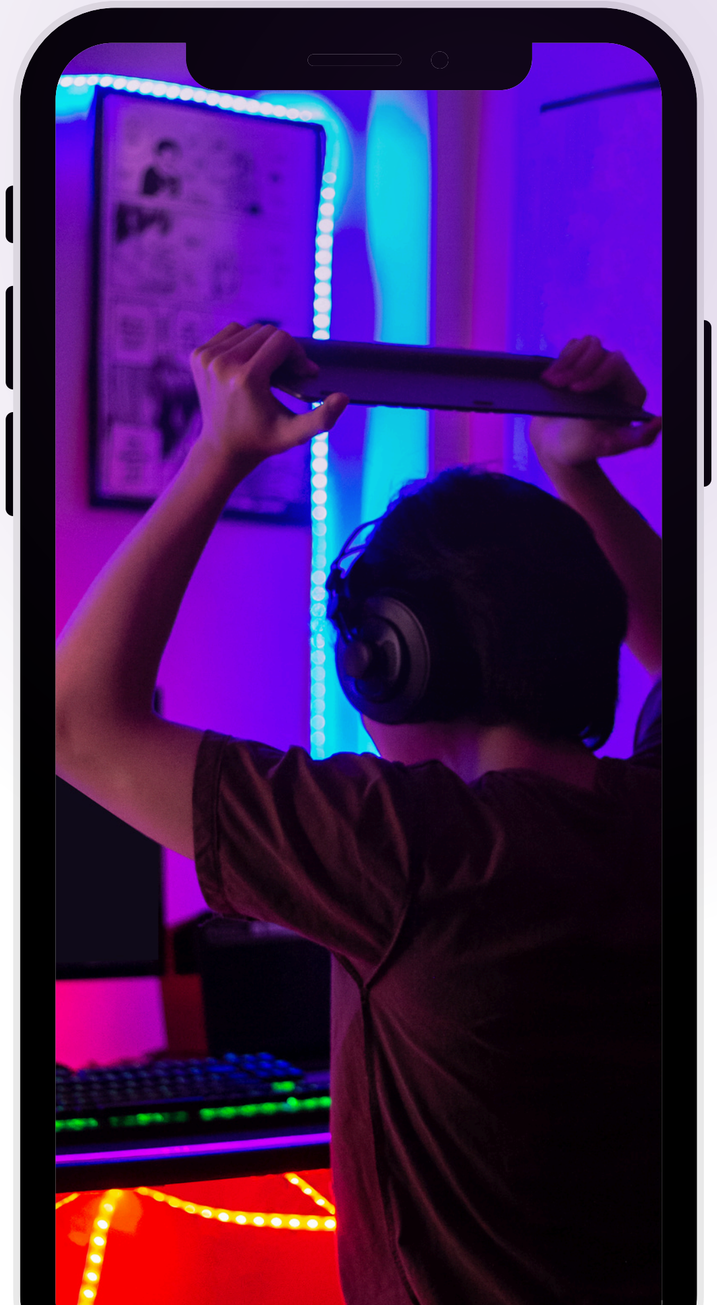
- No example project to reference- all design choices were trial and error.
- defining effective evaluation metrics.

Key Insight:

- The core insight was that precisely defining the reward function is critical. Without a clear and well-designed reward, the agent cannot learn which recommendations are actually better, and overall performance cannot reliably improve.

Future Directions:

- Refine and improve our evaluation metrics for better assessment.



THANK YOU