

Can AI Play Like a Human?

Comparative Evaluation of LLM, DRL, and MCTS for Personality-Driven Playtesting

Ben Stynes

MSc Applied Artificial Intelligence
Supervisor: Libor Zachoval

1. Introduction

- Modern games are complex and difficult to test at scale
- Human playtesting is costly and time-intensive
- AI methods improve efficiency but lack realism

So that means current AI fails to replicate human behavior and personality

2. Research Question

How effectively can LLM-based AI playtesting agents simulate human personality-driven playstyles compared to DRL and MCTS agents in a controlled turn-based environment?

3. State of the Art/literature review

DRL (Deep Reinforcement Learning)

- Learns optimal strategies via rewards
- High performance and adaptability
- Produces predictable behavior
- Limited human-like variability

MCTS (Monte Carlo Tree Search)

- Simulates future decision paths
- Focused on optimization
- Limited personality realism

LLM (Large Language Model)

- Context-aware reasoning
- Behavior influenced by personality
- Supports variability and non-optimal decisions

4. Methodology

Game: Turn-based dungeon crawler (Godot)

Agents: DRL, MCTS, LLM

Tools: Python, GDScript, Hugging Face, PyTorch

- Controlled environment with identical conditions
- LLM uses personality-based prompting
- DRL uses reward optimization
- MCTS uses decision tree simulation
- Evaluation across multiple gameplay sessions

5. Evaluation Metrics

Performance

- Objective completion rate
- Survival duration
- Score efficiency

Behavioral

- Action diversity
- Strategy switching frequency
- Risk-taking patterns

Computational

- Decision time per turn
- Memory usage
- Processing cost

6. Early Indications and future

- LLMs expected to show more human-like behavior
 - DRL/MCTS expected to be more efficient but less realistic
- Next Steps:**
- Implement agents
 - Pilot testing
 - Data collection
 - Behavioral analysis