Project Two Design Defense

In solving the treasure hunt maze, there are notable differences between how a human and an artificial intelligence (AI) agent approach the problem. “Artificial intelligence is the most powerful tool humans have ever invented.” (OpenAI 2025) A human player would rely on cognitive mapping, memory, and logical deduction to find the shortest path to the treasure while avoiding obstacles. They might try different paths, backtrack when necessary, and remember previous mistakes to optimize their approach. In contrast, the AI agent does not have innate problem-solving intuition; instead, it must learn from interactions with the environment. Using reinforcement learning, particularly deep Q-learning, the pirate agent navigates the maze by associating rewards and penalties with different actions, gradually improving its ability to reach the goal efficiently.

The intelligent agent follows a structured process to find the treasure. Initially, it explores the maze using both random moves (exploration) and learned knowledge (exploitation). Each action updates its understanding of the environment by storing experiences and adjusting its neural network weights. Through repeated trials, the agent refines its decision-making process, prioritizing paths that lead to higher cumulative rewards. This approach mimics human learning but is purely data-driven rather than relying on intuition. A key similarity between human and machine approaches is the reliance on trial and error. However, a major difference is that the AI can process thousands of iterations rapidly, whereas humans rely on limited memory and strategic foresight.

The balance between exploration and exploitation is critical in reinforcement learning. Exploration involves taking random actions to discover new strategies, while exploitation leverages past knowledge to maximize rewards. In this pathfinding scenario, an ideal balance might involve a **high initial exploration rate (ε ≈ 1.0)** to map out the environment, followed by a gradual reduction (ε → 0.01) as the agent gains confidence in optimal paths. This ensures the pirate efficiently learns without getting stuck in suboptimal solutions. Reinforcement learning allows the agent to determine the most effective path by continuously updating its policy based on received rewards, reinforcing successful strategies while discarding ineffective ones.

Deep Q-learning was implemented to enable the pirate agent to solve the maze efficiently. Using a neural network, the agent approximates the Q-values of different state-action pairs, allowing it to generalize from past experiences. The model updates its weights through backpropagation, using the Bellman equation to estimate future rewards. Key optimizations included **a training loop with a rolling win-rate window (100 episodes), dynamic loss accumulation, and an ε-greedy policy for balancing exploration and exploitation**. This implementation allows the pirate to learn optimal paths progressively, improving performance over multiple training epochs. By leveraging deep reinforcement learning, the AI effectively solves the complex problem of navigating an uncertain environment with minimal human intervention.

References:

*Introducing the Intelligence Age*. (2025). Openai.com. <https://openai.com/global-affairs/introducing-the-intelligence-age/>

‌ *Mastering Board Games by External and Internal Planning with Language Models*. (2025, February 5). Google DeepMind. <https://deepmind.google/research/publications/139455/>

‌

‌