Project Two Design Defense

Analysis

Humans and machines learn very differently. Humans can conceptualize and think whereas machines learn from a dataset and make decisions based on that dat. Humans form opinions not just from their current circumstances but on connections and memories from their lifetime. Machines need specific algorithms and models to learn from data (Kaushik, 2023). In the case of the pirate’s maze, a human would likely follow along the walls of the maze, using their vision to avoid dead ends. The human would also likely only solve the puzzle one time, while taking wrong turns, and eventually finding the finish. This is not the same way that an intelligent agent would solve the puzzle. The intelligent agent would run through a training loop hundreds or thousands of times, depending on how it was programmed. It would use the feedback to get better each time. Initially, it would likely be learning based on reward. It would solve the puzzle until the appropriate win rate was achieved. Notable similarities are that each learn by trial and error, or in machines, reinforcement learning (Agdestein, 2025). There are numerable differences though. Humans need a minimal amount of trial and error to solve the puzzle. Machines need much more trials to learn. Humans also avoid making the same mistakes, while the agent might repeat old mistakes to learn further from them.

Assessment

Exploration is when the model goes through methods of discovery to find its path. Exploitation is when the model prioritizes its moves based on value. The goal of exploitation is to get the highest reward (GeeksforGeeks, 2024). Both are quite valuable when training a model, so it is best to leverage both. In the case of the pirate maze, it is best to have the model favor exploration in the beginning. This is because the model can then use that knowledge to fuel its exploitation. This will lead to the model finding the solution more quickly. If you favor exploitation too quickly, the model will likely spend too much time gaining rewards and not enough time exploring the environment. Reinforcement learning helps the agent determine the path to the goal by receiving feedback through rewards (Agdestein, 2025). It then uses that feedback to learn through repetitive plays to determine the best path, which in our case would be the shortest.

Evaluation

In our version of the pirate maze, we employed several facets of deep Q-learning with a neural network. We used exploitation to find the best action to make through the if/else epsilon statement. In our statement it either explores by using random actions or exploits by taking a and action based upon rewards. Optimization was done through small batch training with limiting epochs to 10 and batch sizes to 50. Finally, we monitored progress through win rate tracking with hsize and completion check to ensure that the agent was consistently solving the maze.

Resources

Agdestein, I. (2025, February 27). *Reinforcement Learning: AI’s Trial-and-Error Method*. Focalx - Ai Powered Vehicle Inspection. https://focalx.ai/ai/reinforcement-learning-explained/

GeeksforGeeks. (2024, May 18). *Exploitation and Exploration in Machine Learning*. GeeksforGeeks. https://www.geeksforgeeks.org/machine-learning/exploitation-and-exploration-in-machine-learning/

Kaushik, H. (2023, September 21). *Machine Learning vs Human Learning: The Battle for Future Dominance*. Medium. https://medium.com/@himanshubangalore/machine-learning-vs-human-learning-the-battle-for-future-dominance-fa7a1c99cd0c