Design Defense

Humans and artificial intelligence tackle problems completely differently. Humans have the ability to rationalize and think through problems as well as obtain knowledge through various senses. The pirate maze for example would be far easier for a human to solve as they can view which ways they can and cannot go and make decisions based on that information instantly. A human would be able to walk along the walls of the maze and inevitably arrive at the treasure. An intelligent agent cannot solve a maze like this so the best solution would be that of trial and error. The agent can obtain points for walking onto a free space or lose them for walking into a wall or a space traveled before. Eventually the agent would be able to locate the treasure just like the human could, but the agent might take a significantly more convoluted path to reach it.

Exploitation and exploration are two sides of an optimal training algorithm. Exploitation allows the agent to be greedy and take pathways that optimize its point gain with its current set of knowledge. Exploration aims to forego point gain in favor of increasing total knowledge for the future. In other words, exploitation is a short-term gain whereas exploration is a long term one. Both are vital to ensuring the agent is able to eventually solve the maze efficiently, but a good balance would be one that slightly favors exploration as a winding maze is a challenge that naturally requires knowledge. Reinforcement learning helps teach the agent that random movements aren’t beneficial while also reinforcing that thoughtful movement through the maze is greatly beneficial to reaching the treasure.

Deep Q-learning was implemented to allow the agent to better understand the long-term goals of reaching the treasure and allowed for the agent to develop a memory of the maze through neural networks. By decided what the highest Q-value was of a given state, the agent was able to optimize its movements towards the goal.