**Describe the steps a human being would take to solve this maze.**

Ideally, a human would engage in the maze on a visual level, assessing the standout portions of the maze image first, being the darkened squares among the white grid. They would then need to discover the starting point and end point. Once this information is acquired, they would reassess the pathway and find the chain of white grid spaces that remains uninterrupted by a black square all the way to the end point and then follow that path.

For a maze of this size, there will likely be little to no trial and error. Once the human realizes that going in one of the two main directions leads to failure, and the other is the only available path to the endpoint, visualizing the available gap is straightforward.

**Describe the steps your intelligent agent is taking to solve this pathfinding problem.**

The agent starts off initializing in a random free cell within the maze, then observing the current state of the maze. Based on the agent’s position, it discovers valid actions and selects one using an epsilon greedy strategy. The agent explores by using known successful actions from learning Q-values. The agent continues to select a valid action, receives a reward, and stores the experience in memory. Experience replay trains the agent’s neural network by sampling past experience. This helps the agents minimize the difference between predicted and target Q-values.

Through its training, the agent tracks its win rate and adjusts its probability as it becomes more consistently successful. Training will end early if a consistent high win rate is achieved. This means the optimal path to the treasure has been discovered.

**What are the similarities and differences between these two approaches?**

Both the human and the intelligent agent begin the process by assessing their environment in order to identify options and potential actions to make. This is done by understanding the information at their disposal. They each must understand the current state of the maze and avoid obstacles. The human operates on memory and intuition, whereas the agent updates Q-values and uses trial and error. Both seek the most efficient path to the treasure and adjust as their understanding of the maze deepens.

The human and agent processes differ in many ways of course. Humans use visual indicators along with intuition and heuristics. Heuristics, in short, are the mental pathways used to problem solve in a quick yet sufficient manner. (Chen, 2024) For example, humans will often read the beginning and end of words rather than the entirety as this leads to efficient reading. The trade off is a potential lapse in accuracy. If you are proofreading, this natural tendency is now acting against you.

The agent relies on exploration and reward-based learning, whereas humans will operate under the benefits of prior knowledge of mazes and problem solving. Humans are excellent at pattern recognition and adjusting based on that recognition paired with experience. The agent uses a systematic exploration approach with an epsilon-greedy strategy. The human follows a flexible and dynamic approach while the agent uses a gradual data-driven approach.

**What is the difference between exploitation and exploration? What is the ideal proportion of exploitation and exploration for this pathfinding problem? Explain your reasoning.**

Exploitation is the process of using a particular method to achieve success, while exploration is the process of discovering alternative methods of achieving success. (GeeksforGeeks, 2024) To find the ideal proportion, the context of the specific situation and state of the agent must be considered.

If an agent is achieving a high success rate with multiple methods yet continues to look for new options rather than focusing on building upon current successful paths, perhaps the model is over-exploring. (GeeksforGeeks, 2024) Conversely, if a model is prioritizing short term success at the cost of long terms goals, the model may need to be adjusted to do more exploring and less exploiting.

**How can reinforcement learning help to determine the path to the goal (the treasure) by the agent (the pirate)?**

Through reinforcement learning, the agent develops an understanding of the maze environment through interacting with it. The agent will perform actions, receive rewards or penalties as consequence, and then update its strategy based on those results.

The agent does not begin with prior knowledge, it must explore and receive feedback and come to a gradual understanding of the environment. If the agent takes an action that brings it closer to the treasure, it receives positive feedback and if it hits an obstacle, it receives negative feedback. This leads to the agent following similar actions to the ones that achieved positive feedback and avoiding others.

**How did you implement deep Q-learning using neural networks for this game?**

The deep Q-learning was implemented using a neural network to approximate the Q-value function. This function predicts expected rewards for actions. The Keras based neural network includes input, hidden, and output layers where the input takes the maze state, the hidden layers capture patterns, and the output predicts the Q-values for the four available actions.

With the epsilon greedy strategy, the agent trains by exploring the maze and storing its experiences in a pattern of state, action, reward feedback, next state. The network uses the Bellman equation to minimize the difference between Q-values. As our agent learns, epsilon decays, which increases its tendency to exploit the best-known paths. This helps us arrive at an efficient process to get the agent to the treasure.

# References

Chen, J. (2024, May 02). *Heuristics: Definition, Pros & Cons, and Examples*. Retrieved from Investopedia: https://www.investopedia.com/terms/h/heuristics.asp#:~:text=Heuristics%20are%20mental%20shortcuts%20for,up%20analysis%20and%20investment%20decisions.

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