**Project 2**

In the Treasure Hunt game there are several differences and similarities to the way a human and an AI would play the game. In the case of a human playing the Treasure Hunt game, the process a human would take is a very straightforward pattern. A human would read the instructions of the game and determine the best way to proceed to find the treasure. Through trial and error, a human would adjust their approach to improve the way they play the game. Over time a human would improve their score by learning from their past mistakes. On the other hand, our intelligent agent solves the problem in a different way. The intelligent agent learns by reinforced learning. The approach of the intelligent agent is to determine the best score from all the possible moves it can make. The AI stores the information from past games and uses this data to improve its play. After numerous games played the AI will eventually create its most optimal game to discover the treasure. There are numerous similarities between a human and intelligent agent approach. The biggest similarity is they both learn from experience. Using information from past games humans and AI learn from their blunders and will keep improving the more they play the game. Also, they both ultimately want to win with the best approach they know. There are also many differences between the way a human and AI would play. The human will have an advantage in the initial games played versus an AI. This is because the human will draw upon past life experience to realize that it should avoid obstacles, whereas the AI would have to land on the obstacle, and get the negative score to learn to avoid the obstacles through reinforcement. Another difference is the improvement of a human and an AI. The AI will try to get the highest score possible with each move eventually creating the most optimal way to play. A human most of the time will try to improve their own play and strategy and not necessary try to find the most optimal way to play the game. Also, the AI has the advantage of speed and time. An AI can play more games more times than a human.

As it relates to machine learning exploration means that the AI will probe a much larger portion of the spaces in the Treasure Hunt game. The hope is the AI can discover new promising solutions that have not been explored yet by the AI. The AI then analyzes the benefit of the new path through exploiting. In exploiting the AI will refine its path using previous routes to improve its score. The ideal proportion of exploitation and exploration for this pathfinding problem is a small value close to 0.1. This can be determined by running several values for the epsilon and showing when the model obtains 100%-win rate. This value is the exploration factor and gives us the percentage of games the AI will explore a new path instead of refining its old games. At 0.1 the 100%-win rate will occur on the 76 game. At 0.15 the 100%-win rate will occur on the 154 game. At 0.05 the 100%-win rate will occur on the 140 game. This shows that our optimal exploitation and exploration value should be around 0.1.

Reinforcement learning can help to determine the path to the goal by the agent through incentives. By assigning positive or negative scores to each space the AI can receive a feedback with each move. The goal of the AI is to achieve the highest score possible. By playing more and more games the AI learns from its past games and past scores to refine the way it plays. It will move to the space where it will be given the highest score and avoid the spaces that will give it negative scores. This is essentially how the AI learns through reinforcement.

To create the neural network model for our deep Q-learning we needed multiple layers, the activation, optimizer, and loss functions that are used to train the model. In the main function of the code we implemented a deep Q-learning to find the best possible navigation to find the treasure cell with the maximum reward. This is completed by determining the optimal number of epochs to achieve a 100%-win rate. To create the code for the deep Q-learning model for each epoch, we assigned a randomly selected free cell for the agent and then reset the maze with the agent to the original position, using the reset method. Then we created a loop to assign the game state and implemented a while loop to assign action to the agent determined by the exploration and reward function of the code. We then store the actions of the games played using the episode function to train our neural network to evaluate its play.

References:

Janardhanan PS on February 4, 2020 at 8:43pmV. B. (n.d.). *Human Learning and Machine Learning - How they differ?* Data Science Central. https://www.datasciencecentral.com/profiles/blogs/human-learning-and-machine-learning-how-they-differ.

Sergios Karagiannakos. (2018, November 17). *The idea behind Actor-Critics and how A2C and A3C improve them*. AI Summer. https://theaisummer.com/Actor\_critics/.