PROJECT 2- Design Defense

CS370- Emerging trends in Computer Science- AI

Gonzalo Patino

2025

**Differences Between Human and Machine Approaches to Solving Problems**

Humans typically approach a maze using perception, spatial reasoning, and heuristics, such as visually scanning corridors, anticipating dead ends before entering them, and forming a mental model of the layout. This capacity to abstract, predict, and prune options often yields solutions in relatively few attempts. The machine agent, in contrast, begins without prior knowledge, interacts with the grid through repeated trials, and improves its behavior by mapping states and actions to expected returns. In my notebook implementation, the agent learned via deep Q-learning, updating action values from rewards and penalties while gradually improving its policy through replayed experience rather than intuition or visualization.

**Steps a Human Being Would Take to Solve This Maze**

A human solver would begin at the top-left cell, identify permissible moves, then choose a path that appears to progress toward the bottom-right treasure cell specified in the project specifications. As obstacles or dead ends appear, the solver would backtrack, avoid revisiting unproductive branches, and leverage short-term memory to record which options have already failed. This process uses foresight and selective attention, which reduces unnecessary wandering and tends to accelerate convergence on a valid route.

**Steps the Intelligent Agent Takes to Solve the Pathfinding Problem**

The intelligent agent proceeds in structured cycles. At each step, it observes the environment state produced by the Maze class, selects an action using its current policy, receives a reward, and stores the transition tuple in experience replay. Training samples are drawn from memory to update a neural network that estimates Q-values for each action, which implements the Q-learning target update that blends immediate reward with discounted estimates of future return (Sutton & Barto, 2018). Through many episodes, this iterative update reduces the prediction error and leads to a policy that increasingly favors actions that reach the goal.

**Similarities and Differences Between Human and Machine Approaches**

Both approaches require exploration, memory, and feedback. Humans explore by trying new branches, remember failures, and update their plan; accordingly, the agent does the same with stochastic action selection, a replay buffer, and gradient-based updates. Important differences remain. Human reasoning is symbolic and heuristic, which allows jump-ahead prediction and rapid pruning, while the agent’s learning is statistical and incremental. Humans often solve in a handful of attempts, whereas the agent needs many training episodes to converge on a stable policy, although the result can be very reliable once learned.

**Purpose of the Intelligent Agent in Pathfinding**

The agent’s purpose is to learn an autonomous navigation policy that maximizes cumulative reward while respecting the environment’s constraints. In this project, the environment and reward rules are defined by the specifications, including the start in the top-left and the treasure at the bottom-right. The broader objective is to demonstrate how reinforcement learning transforms interaction data into effective behavior without explicit path encodings, a paradigm relevant to robotics, autonomous game agents, and other sequential decision domains outlined in the assignment’s directions and rubric.

**Difference Between Exploitation and Exploration**

Exploration acquires information by choosing unfamiliar actions, while exploitation leverages current knowledge to choose the action with the highest expected return. In my notebook, epsilon was initialized at 0.1, so the agent began with limited exploration and primarily exploited its current Q-estimates. This configuration still allows discovery of alternatives while reducing aimless wandering. Over training, the policy becomes increasingly deterministic as Q-value estimates improve, and the replay process corrects early biases that arise from starting with low exploration (Mnih et al., 2015).

**How Reinforcement Learning Determines the Path to the Goal**

Reinforcement learning solves the path by associating state–action choices with the returns they ultimately produce. Each step yields a reward according to the environment rules: +1.0 for reaching the treasure, negative penalties for invalid or wandering moves, and per-step costs to discourage loops. The Q-learning update integrates these immediate signals with discounted future value, which shifts action preferences toward sequences that reach the goal. Over repeated episodes, the agent’s Q-values reflect the true long-term desirability of actions, and the greedy policy with respect to these values yields a reliable path.

**Use of Algorithms to Solve Complex Problems and Implementation of Deep Q-Learning**

Deep Q-learning scales tabular Q-learning by using a neural network to approximate Q(s, a) across large or continuous state spaces. In this project, the provided replay class stores transitions and constructs supervised targets that embed the Bellman optimality update, then the network parameters are adjusted by minimizing the discrepancy between predictions and these targets. This approach reduces temporal correlation through random sampling, stabilizes training, and improves sample efficiency, which is why it became a standard method for learning control policies directly from high-dimensional observations (Mnih et al., 2015).

**Conclusion**

The final notebook conforms to the assignment specifications and rubric. The agent learns from interaction, respects the movement and reward rules defined for the maze, and converts a small initial exploratory signal (epsilon 0.1) into a high-performing policy using replay and neural value approximation. Juxtaposed with human strategy, the agent’s success illustrates how machine learning replaces intuition with principled estimation, enabling robust, autonomous navigation in a constrained, well-specified environment.

**References**

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., … Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533. https://doi.org/10.1038/nature14236

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.