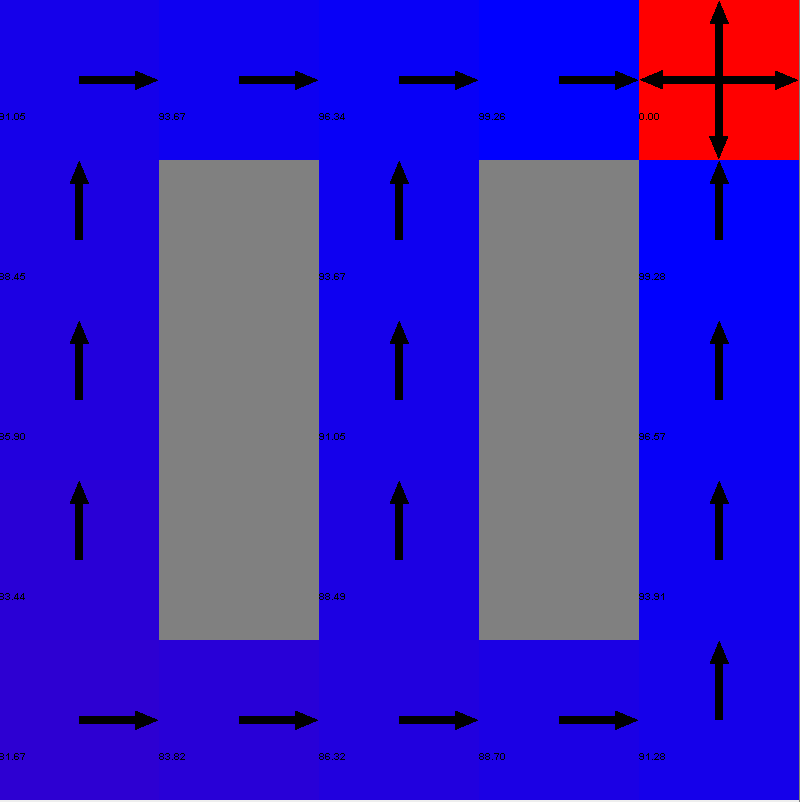
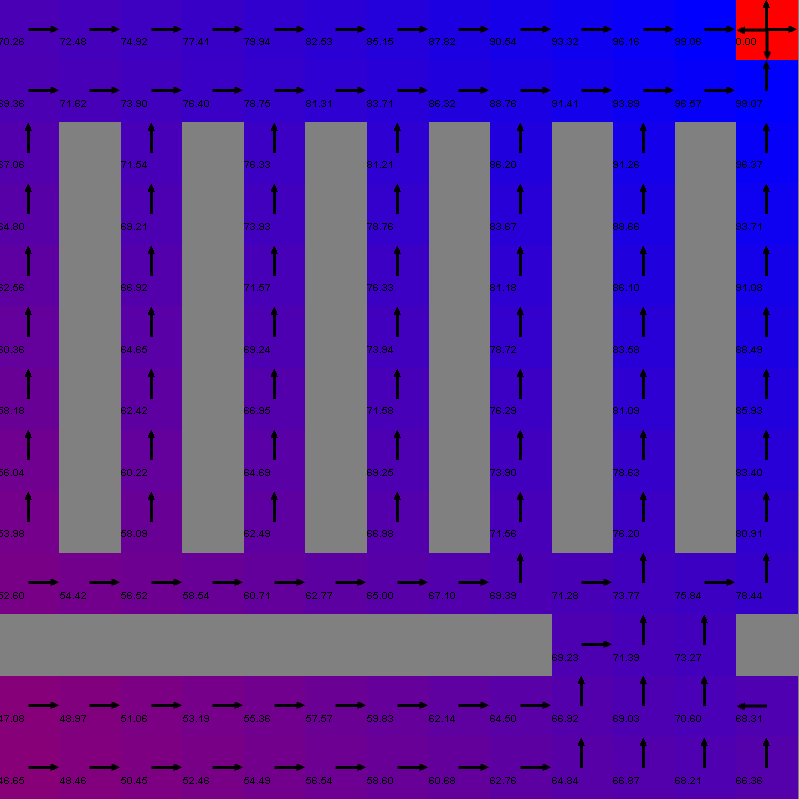
For my MDP problems, I chose two grid world problems, both with some grounding in a real situation. I work for a large manufacturing company and in my prior role, I was working at one of the largest operational warehouses in the world. It’s no secret that automation and machine learning will become a bigger part of our operation in facilities such as those, so I was interested in applying what I learned about reinforcement learning to grid worlds that mimic subsections of our warehouses. Below is my grid world with a “small” number of states.



The rewards and policy shown on the grid above are those calculated by value iteration. This layout is meant to loosely mimic the rows of shelves where parts are stored in the warehouse (the resemblance will be more evident in the “large” state example). Having the ability to automatically determine optimal routes throughout our warehouse can result in significant cost savings, given that we pick millions of parts per year and our buildings and shelving structures are always being moved for optimization purposes.

Below is the “large” state MDP problem layout. The rewards and policy on this visual are that of policy iteration as opposed to value iteration for the small state problem.



In this grid, it’s easier to see the characteristics of a warehouse—starting from the bottom left, the horizontal “wall” could be seen as a wall between different buildings with the gap being the door to get from one to another. The parallel vertical lines in the upper 2/3 of the grid are “shelves” where parts are stored.