**Task 2 Solution**

**Name: Ho Phi Long**

**Matric: A0170751J**

To solve task 2 large stochastic environment, I have created a deep Q learning agent with modification to the reward and goal system taking inspiration from OpenAI team research paper: Hindsight Experience Replay. In order to create a reward system that agent can use to train corresponding behavior to achieve the goal, I create a new goal system and use together with the environment reward to train the agent

* Instead of the main goal to reach the finishing point, agent now used a sub goal in conjunction with the main goal that being generated as followed: each time the game end, agent will take it current position at the end of that game and treat it’s as a sub-goal.
* The full reward will be given if agent can reach that sub goal and agent will try to optimize reaching that sub-goal. The aim of this is to help agent to realize patterns like how to avoid cars using the observation and using appropriate action that can help to complete this task: how to get to a certain position in the environment.
* After finish training this agent, the sub-goal system is taken away from the agent and now use the same knowledge to try optimizing the main goal of reaching the finishing point
* To ensure agent can work well with horizon of 40, 20% of rewards is also added to agent when agent manage to reach goal before step 40 for both the sub goal and the main goal

To track the training process, average steps when terminate, average distance from goal and average rewards is used to judge the efficiency and halt the training if necessary.