

Another approach is to use Monte Carlo Tree Search as a baseline to compare against the other two, where by using a score system to determine who is most likely to be werewolf, we can create a rough belief distribution on who is more likely to be werewolf, and during the tree search we can randomly sample a game state out of the distribution for playout strategy.

When compared with vanilla AI the Monte Carlo is able to achieve an overall win rate of 47%, so the good news is that it's able to run and work, but the result is not impressive because the implementation is very naive when talking about the complex game such as werewolf, the playout strategy could be improved to be more realistic, instead of the current implementation where I assume the villager will always vote out the current werewolf.

The score system is also not great and easily exploitable, in that I assume the werewolf will be the one that's getting accused of being a werewolf, but in a social deduction game that's rarely the case.

The original plan is to use a neural network to evaluate the game state instead of a score system, but right now it still needs tweaking and is not in a state where it can be used, because it's a lot more complicated than I initially assumed, where more time is needed.