-What was actually done

For human-being to learn how machines are learning using neural networking so we can understand better of the pattern. So we built a machine to learn about how other machine is working, how they are observing the world and understanding the world by observing them, so that we are able to know how they’re processing the existing data and material and logic behind it to build, the main purpose of all of these is to understand better of what will make machine go wrong in order for us to improve them better in the future and make more perfect generations of them to work better for human.

-What was the experiments (with numbers and results)

5 experiments were done in order to actually “understand” the machine. By saying “understand”, i mean with existing information we could predict the behavior of machine and the reason behind it so that we could avoid it in the future design.

1st, simple but random intelligent identities for the ToMnet to learn, to get the characteristics of approximating Bayesoptimal hierarchical inference.

2nd, show ToMnet the targets of algorithm intelligent identities, the most importantly about how they balance cost and rewards.

3rd, show that the ToMnet learns to from different types of deep reinforcement learning intelligent identities. So it can capture the key component of variations and form the abstract embeddings. “Also showed ToMnet can discover new abstractions about the space of behaviour.”

4th, “we show that when the ToMnet is trained on deep RL agents acting in POMDPs, it implicitly learns that these agents can hold false beliefs about the world, this is also a core component of humans’ Theory of Mind.”

5th, “we show that the ToMnet can be trained to predict agents’ belief states as well, revealing agents’ false beliefs explicitly. also show that the ToMnet can infer what different agents are able to see, and what they therefore will tend to believe, from their behaviour alone.”

-What ideas from ToM were used?

This is to build up the theory of mindset structure with the basics elements from faith, belief, desire, emotions and knowledge.

-What is the solution architecture?

2 inputs inbuilt to one network. By a intermediate system, we reduce the dimensions off the space to understand the machine inner logics with an easier way rather than modifying the internal structure for human to read. ToMnet is more like an lacked interface between human and machines in this way