In the article “Artificial Intelligence has learned to probe the minds of other computers”, Matthew Hutson talks about an attempt at finding out what one intelligent computer system could learn about others, so that we, humans, could understand them better. Humans are not able to comprehend the complex algorithms (*machine learning has implemented*) in these intelligent computer systems, and that's why scientists conducted an experiment to find out what one computer could find out about how other computers work.
A research scientist named Neil Rabinowitz, together with his colleagues, was able to create a ‘thinking’ computer system called “MNet”, which roughly simulated the human brain, to learn more about other computers, learn about their ways and even predict their next steps.

The experiment was set in a virtual room, where the computers under observation were moving around collecting colored boxes to get points. After some training the machine that observed them from a room above, could spot the different “species” of character, that the computers were split into, after just a few steps of observing them. One “species” could not see, another one could not remember its recent steps and the third one could see and remember its steps both. The intelligent computer system could not only do that, but it also predicted what steps each “species” would most probably take next. With a final test they also figured out that the theory of mind computer system could also understand when one of the computers observed held a false belief.

To create this intelligent machine, ideas from ToM were used. For example, the computer is able to learn from experiences, just as humans are doing from the moment they are born. It learns the other computers tendencies by observing and understanding what they have done in the past, forms an understanding of their beliefs and puts all of this into context to predict future actions. The computer can do all of this, by basing it on what it knows about itself. Human brains function the same way to understand and interact with others around them. The computer has shown many ways of how it can operate very similar to a human brain, by learning experiences, using the knowledge it has about itself to create assumptions about others and learning skills on their own by being able to do all of these things. In the future these intelligent ‘thinking’ computers might even be able to decept people, by understanding what false beliefs are and persuading humans of those.

A computer scientist and psychologist called Josh Tenenbaum had also worked on creating a computational model in a way that implements capacities of theory of mind. Different to Rabinowitz’s neural networks approach, Tenenbaum’s system is based on a form of probabilistic reasoning. Although going down the neural networks path might be more efficient, it is lacking the capability to adapt in new environments, because it is attached to the contexts in which it’s been trained in. Tenenbaum’s system on the other hand is able to adapt in various environments. He says the solution might be to combine both of these approaches to create something even more powerful.