Summary: Minds, Brains and Computers by John Searle

18 October 2019 - Paul Meierling (6.844)

Overall summary of the paper

John Searle argues about whether or not a computer with the right kind of program would be able to understand and think. His opinion is that this is not the case. His argument is based on the analogy of a person in a room producing answers to chinese questions without understanding chinese.

Structure of the paper

Searle differentiates between two types of AI. Weak AI, which sees the computer merely as a tool and strong AI, which describes a program/computer that understands and is able to reason. He uses the work of Roger Schank as an example, who developed a machine that is able to produce answers based on a story, which did only mentioned those answers implicitly (e.g. did a man eat a hamburger if he left the restaurant with a full stomach). According to Searle this does not constitute as understanding and he makes his point by using the analogy of a man locked in a room. That man, who does not understand Chinese, is being given multiple sets of Chinese symbols and a set of rules in a language that he does understand. These rules show how the symbols are related to each other and allow the person to to produce answers to Chinese questions without actually understanding Chinese. However another person standing outside the room would think the person in the room is able to understand the questions as well as the language. This analogy can be applied to questions is any kind of “human language”, which are read by a computer who then appears to understand the language, but actually only does some kind of pattern matching. Searle further argues that we often ascribe human understanding to machines even though they simple relate input and output (e.g. the thermostat knows when to change the temperature). He then argues that while machines can think (humans are just some form of machines), instantiating a program itself is not a sufficient condition of understanding. Firstly because formal manipulation of symbols does not contain intention, which programs only have through the programmers who wrote them. Secondly, programs are by their nature purley formal, whereas intentions are not. Thirdly, thoughts are the subproduct of our brain, whereas a program is not a product of a computer. A simulation of understanding would therefore not lead to actual understanding.

Discussion and implications

While the argument about Roger Shanks paper is easy to understand, Searle does not explore the meaning of a “program” or the human way of understanding. In my opinion a hypothetical program could be built that contains an exact copy of my brain thus allowing it to “understand” in a way we humans do. As of know we do not know if there is a deeper meaning in our understanding or if we are also just pattern matching input to output, but in a much more complex way. Building a program that is able to take in all sensory inputs from a human and give the same outputs as a human would in a way indistinguishable from a human. How we get from input to output in a way does not matter as much if the output is in all cases the same. E.g. this hypothetical program, would think about everything the same as I do and also act the same way I do. Which in my opinion is the same as understanding as well as I do.