There is certainly a lot of excitement throughout the tech community about the promise of artificial intelligence or AI, as it is more commonly known. And while many of the advances are impressive compared to where computer science was only a decade ago there is a lot more hype than fact in many of the more outlandish claims being made. Skynet from the Terminator Movies or the Machines from the Matrix are not soon to take over nor are they likely to do so for many generations. However, there is a distinct possibility that AI may actually be able to make competent decisions in the near future but only if the community takes a broader focus than it is apparently taking right now.

Now some may object that AI is enabling all sorts of important activities that would not otherwise work. Reports of scientific discoveries, business approaches, and computational improvements abound so why can’t one conclude that AI is making competent decisions now? The crux of the matter is the definition of what AI is/does and what competent means. To flesh out this argument, let’s take a small regression to discuss what is commonly meant by AI and what is capable of doing now.

Many modern books on AI and a spate of YouTube videos extol the recent advances in AI with particular attention on algorithms like convolutional neural nets and the vast improvements they offer in image classification or on support vector machines or K++ Means algorithms and the power they offer in clustering data. For example, one of my favorite videos on the subject is *But what is a Neural Network?* By 3Blue1Brown

<iframe width="560" height="315" src="https://www.youtube.com/embed/aircAruvnKk" frameborder="0" allow="accelerometer; autoplay; encrypted-media; gyroscope; picture-in-picture" allowfullscreen></iframe>

Grant Sanderson (the voice and vision behind 3Blue1Brown) starts off his video by discussing the remarkable functioning of the human visual cortex that can recognize all sorts of different renderings of the number 3. For example, each of the following glyphs show the string “3” printed in a different font.

<>

Most people (and their remarkable visual system) can tell that each character is a different rendering of the same number. And, as Grant details in his video, convolutional neural networks seem to be able to perform the same recognition.

He moves onto discussing how a neural net can be used to encode similar kinds of pattern recognition for a machine allowing it to recognize edges, or loops, and so and that within its multiple layers can be found the ability to report, with very highly levels of certainty that each of render corresponds to the same underlying digit.

This is quite a step forward in machine vision but does it really constitute artificial intelligence or decision making? Sure, the algorithms can comb through vast amounts of data looking for a prescribed pattern and can ‘decide’ when they have found a candidate. And these results should not be surprising because, after all, neural nets were designed to mimic the human cortex and who is to say that the training the net receives and the way it decomposes images into parts doesn’t mimic what is done in the brain.

Despite those arguments, it is philosophically hard to say that the AI can even come close to making competent decisions.

There are two reason for this assertion, one technical and one philosophical. On the technical front the best operative definition of artificial intelligence is that of the rational agent taken from *Artificial Intelligence: A Modern Approach* by Russell and Norvig. Under this definition, the machine must not only recognize the desired pattern (e.g. deciding that it sees a “3”) but it also has to perform an appropriate action based on that recognition.

To understand how a rational agent takes the appropriate action, Russell and Norvig assert that a rational agent received stimulus, called a percept, (in the case above the pixel values of an image of the rendering of “3”) and then acts so as to maximize the value of some performance metric based on the percept, the sequence of percepts up to that point (i.e. some notion of memory), it’s knowledge about its environment, and the rules that it has for actuating whatever actions it can take.

Simply pattern matching and ‘deciding’ that the pattern is either seen or not doesn’t really meet the definition of a rational agent. For example, no sequence of previous percepts (excepting the initial training) is used by the machine learning techniques currently being enthusiastically pursued. The current systems aren’t capable of continuous adjustment as situations change. To pattern match a “3” the system needs to find two half-loops stack one upon the other. If gradually, the adoption of other styles, say the Roman numeral III, became popular in the percept stream, the net would be stymied.

In addition, it is difficult to call a binary sorting a ‘real decision’. Certainly, it is useful to have such a sorting algorithm that can look at vast amounts of ‘noisy’ data and point out the parts that are of most interest but the judgement of what is of interest still sits with the human element. And, to be sure, this is an important step forward, but it doesn’t really constitute ‘learning’ or ‘rational’ in the human sense.

And this brings up the second point. Traditional philosophy recognizes Three Acts of the Mind. Roughly the three acts break down as follows (using our “3” once again). First Act: the system recognizes the “3” in a sentence. Second Act: the system recognizes the meaning of the sentence “Start the process at 3.” Third Act: the system can reason through chaining of statements to answer “Why didn’t the process start at 3?”. At best, what has been accomplished corresponds to the initial, baby steps into the First Act.

For a system to really exhibit some semblance of rationality it must mimic the three acts, there need to be a hierarchy of different types of agents all working together. A possible example of this would be a system as pictured below:

At the bottom would be some set of agents to perform the First Act, perhaps two differently trained convolutional neural nets or a convolutional neural net and a support vector machine or whatever other combination one could imagine. This layer, like the others, should be tool agnostic, focusing on what should be done not how. The second layer might have an expert system to interpret the percepts from the lower layer within the context that the system finds itself (effectively answering Russell’s and Norvig’s requirement that the rational agent know its percept history and its environment). This layer should also have some way of swapping the tools in the lowest layer or adjusting their operating parameters, allowing it to change how it looks at a problem. Perhaps an Analytical Hierarchy Process or a Genetic Algorithm can be employed to weight the performance of the tools. Finally, in the third layer should be some set of tools for chaining the results from the second layer so that rational decisions could be made. Here the tool set is far more speculative. Perhaps a different kind of expert system or AHP or, perhaps, an A\* algorithm could be used. It really doesn’t make what the tools are but rather what they do. This seems to be the only blueprint for achieving a real AI.