GOLDSMITHS, UNIVERSITY OF LONODON

Do Learning Machines Extend Our Minds?

Juan Pablo de la Vega Castañeda
33370651

Critical Theory
CU71007A
28th April 2015

Do Learning Machines Extend Our Minds?

JUAN PABLO DE LA VEGA CASTAÑEDA
Critical Theory, *Goldsmiths College*
28th April 2015

Introduction

The mind has always been in the centre of philosophical debates. Understanding how we think and which processes allow us to perform as humans is a question that has been asked by philosophers and scientists alike for centuries. Nowadays, in light of recent and accelerating research in fields such as neurology, cognitive science and artificial intelligence, and psychology, to name a few, this debate is livelier than ever.

In a very broad sense, we could think of the mind as something that happens *inside* the body or something that could happen *outside* it; I will focus on the latter because they are the only ones which can incorporate our interactions with the environment as part of our mental processes. In particular, I will focus on Andy Clark's theory of the *extended mind*. The main goal is to build a map which will help put in context the role of the notions of the mind and their influence and applicability in more resent research, specifically in artificial intelligence, especially the field of machine learning, and to investigate whether this new technologies will in fact help us extend our minds or are rather independent creations that compete for the thus far hegemonic place that the brain (not necessarily human) has had as the sole producer of mind.

A very brief historical account of computer science will be presented to understand how the notion of artificial intelligence was conceived, which were its original goals and how machine learning was derived from it. This will also put in perspective, and aid in understanding, the main paradigms that drive current machine learning algorithms and their goals. This account will also help as the building blocks for the argument to determine whether machine learning can be considered as an extension of our minds.

The Mind

The mind has long been point of discussion in philosophy and one of the most influential perspectives is the mind-body dualism as presented by Descartes, known as substance dualism. Although, as Damasio (2004, p.187) is keen to point out, "substance dualism, is no longer mainstream in science or philosophy" it is worth presenting here since it "is probably the view that most human beings today would regard as their own." I will briefly present the theory and counterarguments to it, followed by some alternatives. The focus will be mostly on *externalism* and *functionalism* to construct a path toward the philosophical foundations of the extended mind.

The formulation, as described by Ryle (1973, p.13-14) goes as follows: "every human being is both a body and a mind. His body and his mind are ordinarily harnessed together…Human bodies are in space and are subject to the mechanical laws which govern all other bodies in space…minds are not in space, nor are their operations subject to mechanical laws." who emphatically disagrees with it saying that "it [body-mind dualism] is entirely false…not only in detail but in principle"(ibid., p.17). He argues that it is a "category-mistake" which consists on being "competent to apply concepts…but are still liable in their abstract thinking to allocate those concepts to logical types to which they do not belong"(ibid., p.17) and particularly in the case of the mind and the body, "the believe that they are terms of the same logical type."(ibid., p.23) His position is a functionalist one where

> [t]he statement 'the mind is its own place', as theorists might construe it, is not true, for the mind is not even a metaphorical 'place'. On the contrary, the chessboard, the platform, the scholar's desk, the judge's bench, the lorry-driver's seat, the studio and the football field are among its places. These are where people work and play stupidly or intelligently. 'Mind'…is not the name of another tool with which work is done (ibid., p.50)

and it is made more clear in the following statement: "the styles and procedures of peopleś activities *are* the way their minds work and are not merely imperfect reflections of the postulated secret processes which were supposed to be the workings of minds."

Another argument against dualism can be given through discoveries in Physics. Braddon-Mitchell and Jackson (1997, p.3-28) briefly discuss different types of dualism and offer a solution through a particular kind of materialism called *physicalism* by arguing

that

> [i]f the mental supervenes on the physical, the physical way things are makes true the psychological way things are in the sense that a certain sentence (or proposition) about the physical way things are entails each and every sentence (or proposition) about the psychological way things are.

Searle (1984, p.17) recognises that we must be careful when discussing the mind-body problem because "consciousness, intentionality, subjectivity and mental causation... are all features of our mental lives" and "any satisfactory account of the mind and of mind-body relations must take account of all four features"(ibid.). Instead, he proposes *biological naturalism* in which "all mental phenomena whether conscious or unconscious, visual or auditory, pains, tickles, itches, thoughts, indeed, all of our mental life, are caused by processes going on in the brain"(ibid., p.18), and later insists that "people actually think, and thinking goes on in their brains... there are all sorts of things going on in the brain at the neurophysiological level that actually cause our thought processes." A harsh criticism of this theory is given by Putnam (1981), who calls this model "brains in a vat" and disproves it by doing "an investigation... of the preconditions of reference and hence of thought – preconditions built in to the nature of our minds themselves."Putnam (ibid., p.16) and concludes that "meanings just aren't in the head"Putnam (ibid., p.19).

Mind being more than the brain is also argued for by Damasio (2004). In his view,

> [t]he mind arises from or in a brain situated within a body-proper with which it interacts; that due to meditation of the brain, the mind is grounded in the body-proper; that the mind has prevailed in evolution because it helps maintain the body-proper; and that the mind arises from or in biological tissue –nerve cells– that share the same characteristics that define other living tissues in the body-proper(ibid., p.191).

He therefore includes the body as part of the mind "In other words, body, brain, and mind are manifestations of a single organism"(ibid., p.195). This notion of mind brings us closer to a useful one for the extended mind theory but we still need to stretch it some more. Noë (2009, p.xii-xiii) on the preface points us in the right direction saying that "our problem is that we have been looking for consciousness where it isn't... We are not locked up in a prison of our own ideas and sensations. The phenomemon [sic] of consciousness... is a world-involving dynamic process." Although in principle this does not seem to depart

from Damasio, since he also agrees that "the mind is a process, not a thing,"(Damasio 2004, p.183) if we follow Noe's argument we will soon realise that this is not the case. Whereas Damasio keeps the mind within the bounds of the body, Noe does not hesitate to incorporate the environment into the thinking process, for him "[w]hat governs the character of our experience– what makes experience the kind of experience it is– is not the neural activity in our brains on its own; it is, rather, our ongoing dynamic relation to objects, a relation that, as in this case, clearly depends on our neural responsiveness to changes in our relation to things"(Noë 2009, p.50). Moreover, tools are also part of of ourselves in his view "what makes them me, what makes them part of my body, is the way my actions take them up. An insofar as I act and feel with my extended body, my mind is extended too."(ibid., p.80)

We have now found philosophical arguments that support the thesis that the mind is not necessarily inside of our brains but it is also possible to conceive of relationships with external objects that can be coupled to our mental processing and thus form part of the mind, at least for the time that they are coupled. This is the argument presented by Clark and D. J. Chalmers (1998) which will be analysed in the next section.

Extended Mind

I will now present the concept of *extended mind* originally introduced by Clark and Chalmers. Some arguments will be given both in favour and against this model of the mind. This will set the groundwork for determining whether it is possible to consider machine learning an extension of our minds.

Taking from the functionalist perspective the notion of multiple realizability, that is "the idea that these [mental] roles could be filled or occupied by quite different kinds of things in different cases"(Braddon-Mitchell and Jackson 1997, p.42), extended mind aims to provide an explanation of the mind which incorporates tools in the environment as part of the mental process. However, whereas functionalism assumes that mental states are internal (ibid., p.41), Clark and Chalmers pursue what they call "*active externalism*, based on the active role of the environment in driving cognitive processes" (Clark and D. Chalmers 2010, p.27). A very clear and concrete description of the extended mind is given by David Chalmers in his foreword to Clark (2008, p.x): "[t]his is the thesis of the extended mind: when parts of the environment are coupled to the brain in the right way, they become parts of the mind."

Clark and Chalmers start their argument by presenting a thought exercise whereby you should think of a person playing Tetris. There are at least two different strategies in the game, namely to rotate the pieces mentally and then decide where the falling block best fits or to rotate them with the press of a button to help you decide where to land the block. The first one is evidently a mental process whereas it would usually be argued that the second one is not. They propose a third option, one in which a person has an implant that allows for the same visualisation of the rotation of the blocks inside one's head, rather than the monitor, via an implanted chip. This, they argue, is clearly cognitive (Clark and D. Chalmers 2010, p.41-42). Through this example they establish the *parity principle* which states that "If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as a part of the cognitive process, then that part of the world is (for that time) part of the cognitive process"(Clark 2010, p.44). It follows that the three examples would be identically regarded as being mental processes.

A second argument is given, again in the form of a thought experiment. They present Inga, who when hearing about an exhibition at the Museum of Modern Art (MoMA) "she thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum."(Clark and D. Chalmers 2010, p.33) In contrast, Otto suffers from Alzheimer's disease and must use a notebook, which he carries with him at all times, to write down any new information that he learns and which he consults to access old information. He, as Inga, learns about the exhibition at the MoMA and, upon deciding to go see it, "[h]e consults the notebook, which says that the museum is on 53rd Street, so he walks to 53rd Street and goes into the museum."(ibid.) This example is used to define the conditions that need to be met for an external resource to be considered part of the cognitive process (Clark 2010, p.59):

1. That a resource be reliably available and typically invoked. (Otto always carries the notebook and won't answer that he "doesn't know" until after he has consulted it.)
2. That any information thus retrieved be more or less automatically endorsed. It should not usually be subject to critical scrutiny (e.g., unlike the opinions of other people). It should be deemed about as trustworthy as something retrieved clearly from biological memory.

3. That information contained in the resource should be easily accessible as and when required.
4. That the information in the notebook has been consciously endorsed at some point in the past and indeed is there as a consequence of this endorsement.

Further clarification on the topic is given by Clark (2008, p.96) on the issue where he emphasises that "[i]t is the way the information is poised to guide reasoning and behavior that counts…It is coarse systemic role that matters, not brute similarities in public behavior [sic]" and an interview with Nobel prize winner Richard Feynman, conducted by historian Charles Weiner, is presented in support of this thesis (Weiner 1973):

> *Feynman*: I actually did the work on the paper…
> *Weiner*: Well, the work was done in your head but the record of it is still here.
> *Feynman*: No, it's not a record, not really, it's working. You have to work on paper and this is the paper. OK?

There are also those who argue against the concept of the extended mind, the most noteworthy is the work of Adams and Aizawa (2001) and it is further argued in Adams and Aizawa (2010).

Adams and Aizawa believe that a rather common mistake people who defend the extended mind do is the "coupling-constitution fallacy"(ibid., p.67) which consists in "draw[ing] attention to cases…in which some object or process is coupled in some fashion to some cognitive agent" and the reaching the conclusion "that the object or process constitutes part of the agent's cognitive apparatus," directly attacking the example of Otto's notebook. One could argue against this, as Menary (2010, p.21) does in his introduction to the book that "the extended mind is not simply an embodied-embedded thesis that treats external props and tools as causally relevant features of the environment. It is a thesis that takes the bodily manipulation of external vehicles as constitutive of cognitive processes." Other arguments they make include the different ways in which the brain processes information as opposed to when there exist "brain-tool combinations" (Adams and Aizawa 2010, p.75) and the distinction between representations derived from the physical world and nonderived content involved in cognitive states(ibid., p.73).

As you can probably begin to suspect this arguments go back and forth without any

clear winner. Nevertheless, this discussion will serve as context and point of departure for the discussion on whether machine learning can be viewed as a form of extended mind.

About computers

Humans have always searched for tools to assist them in everyday activities. From the beginning of history, we have managed to develop tools that aid us in specific tasks, be it a hammer to pound rocks with or a knife to skin an animal. However, this tools did not appear in their present form, but have been a product of the evolution and adaptation to different needs —it is not hard to see the connection between a shovel and a hydraulic excavator—. Tools for intellectual endeavours are no more recent —the first form of the abacus, the counting board, can be traced to around 500 B.C. (Fernandes 2015)— and have been subject to similar evolutionary processes, the latest state of which can be found in the form of machine learning. This section will be dedicated to a brief history of the computational to better understand the position that machine learning occupies in the current context.

Possibly the first mechanical assist is the Pascaline, named after its inventor the French mathematician Blaise Pascal somewhere arround 1642–1644, which was an arithmetic machine to be used for tax calculations (Freiberger and Swaine 2015) and was closely followed by Leibniz's Step Reckoner presented to the Royal Society in 1671 (Belaval 2015). Both of these machines were designed to help in calculations of additions, subtractions multiplications and divisions so that they could be accomplished in less time and with greater precision. This was for a very long time the main goal of the developments in that field until in 1936 Alan Turing presented a paper in response to answer the *Entscheidungsproblem* posed by David Hilbert, "namely the question of whether there exists a definite method which, at least in principle, can be applied to a given proposition to decide whether that proposition is provable." In his paper *On Computable Numbers, with an Application to the* Entscheidungsproblem, A. M. Turing (1937) presents the notion of computation and "computable numbers" as those that "may be described briefly as the real numbers whose expressions as a decimal are calculable by finite means" and then proposes "computable machines" which can perform exactly this kinds of calculations:

> We may compare a man in the process of computing a real number to a machine which is only capable of a finite number of conditions $q_1, q_2, \ldots, q_k$

> which will be called '*m*-configurations'. The machine is supplied with a 'tape' (the analogue of paper) running through it, and divided into sections (called 'squares') each capable of bearing a 'symbol'. At any moment there is just one square... which is 'in the machine'. We may call this square the 'scanned square'. The symbol on the scanned square may be called the 'scanned symbol'. The 'scanned symbol' is the only one of which the machine is, so to speak, 'directly aware'(A. M. Turing 1937, p.231)

He goes on to assert that it is possible to represent infinite numbers by machines which can only understand a finite set of symbols and uses this to prove that a general solution to Hilbert's *Entscheidungsproblem* does not exist. These machines would be later known as *Turnig Machines*. I will not get into further detail but it is important to know that this paper inspired many a scientist and is said to mark the beginning of computer science as we know it today. One of the most relevant concepts developed after this paper is the one developed by John von Neumann after working on the EDAC machine (Neumann 1993). Presented in 1945 after, it the sets the ground rules for of the digital computer architecture, which even today is the basic design. In a nutshell, it conceives of a computer as having three components, namely a control unit, an arithmetic/logic unit and a memory unit, and being able to receive input and produce output. Any machine that follows this design is now known as a *Von Neumann machine*.

In 1950 Alan Turing published another highly influential paper whose goal was to determine a way of knowing whether a machine can actually be intelligent. They way he solves it is by proposing a test for the machine which he called *the imitation game*. "It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart front the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman" (Alan M. Turing 1950). Then he modifies the question to "What will happen when a machine takes the part of A in this game?" and this is now called the *Turing Test*. This single question has given rise to an extensive discussion on whether computers can be intelligent and whether it is possible for them to have a mind, nonetheless this is not within the scope of this essay (for extensive discussion on the topic, see (Sloman 1978; Searle 1984; Miłkowski 2013)). I will focus on another consequence of this question: artificial intelligence as a field of study that aims to provide a computational model of the mind and, in particular on machine learning techniques. This in the hopes of investigating

whether it is possible to extend our minds —in Clark's sense— through machine learning.

One more important paper to bear in mind before entering the field of artificial intelligence is the one presented by Shannon (1948). In this report for Bell Labs, Shannon gives a thorough description of a mathematical model by which information can be represented. Five elements are needed in a communication system:

1. An *information source* which produces a message or sequence of messages to be communicated to the receiving terminal...
2. A *transmitter* which operates on the message in some way to produce a signal suitable for transmission over the channel...
3. The *channel* is merely the medium used to transmit the signal from transmitter to receiver.
4. The *receiver* ordinarily performs the inverse operation of that done by the transmitter, reconstructing the message from the signal.
5. The *destination* is the person (or thing) for whom the message is intended.

This elements can be easily observed in Figure 1

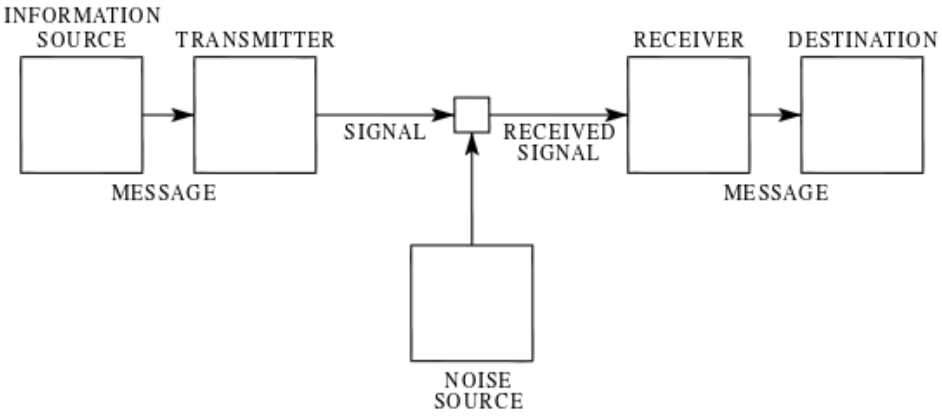

Figure 1: Schematic diagram of a general communication system.

This model, along with the formalisation of information that gave the possibility of represent it in abstract terms, helped set the foundations of a field of study whose goal was tu provide an explanation through computing of how the mind functions. This field was later to be named artificial intelligence.

Artificial Intelligence and Machine Learning

Artificial intelligence is a term that was coined in 1955 in a proposal for a research project and has been present in the computer science community ever since. The proposal includes an analysis of the areas of interest in artificial intelligence (AI) which is worth quoting at length because it defined its focus for years to come (McCarthy et al. 1955):

> We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer…
>
> The following are some aspects of the artificial intelligence problem:
>
> 1. **Automatic Computers** If a machine can do a job, then an automatic calculator can be programmed to simulate the machine…
>
> 2. **How Can a Computer be Programmed to Use a Language** It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture…
>
> 3. **Neuron Nets** How can a set of (hypothetical) neurons be arranged so as to form concepts…
>
> 4. **Theory of the Size of a Calculation** If we are given a well-defined problem (one for which it is possible to test mechanically whether or not a proposed answer is a valid answer) one way of solving it is to try all possible answers in order…it is necessary to have on hand a method of measuring the complexity of calculating devices…[and] a theory of the complexity of functions…
>
> 5. **Self-improvement** Probably a truly intelligent machine will carry out activities which may best be described as self-improvement…
>
> 6. **Abstractions** …[An] attempt to classify these [types of abstractions] and

to describe machine methods of forming abstractions from sensory and other data would seem worthwhile...

7. **Randomness and Creativity** A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of a some [sic] randomness.

If we take a moment to analyse each of these areas we can realise that most of them —if not all— are still alive and well, at least in the conceptual form. For example, we can equate "How Can a Computer be Programmed to Use a Language" with natural language processing (NLP), an area of AI so important nowadays that companies such as Microsoft and Google have dedicated research areas on the topic; [1] if we take "Neuron Nets" it is easy to see the resemblance with artificial neural networks which hold a esteemed position in AI research; and "Theory of the Size of a Calculation" which inspired a mathematical analysis of computational algorithms, known as complexity theory. Many other subsets of artificial intelligence can be described, but I believe these are enough to realise that AI is, and has been from its beginning, a very broad area of study so instead of trying to scrutinise every detail of it, the focus will be in a specific branch called *machine learning*.

Simply put, "machine learning is concerned with the question of how to construct computer programs that improve their performance at some task through experience."(Mitchell 1997, p.17) Although the specific algorithms and mathematical models can vary much more, there are three main strategies, broadly speaking, currently explored in this field: *supervised learning*, *unsupervised learning*, and *inductive learning*.

Supervised learning algorithms generally use a know set of data which includes the corresponding input and response to each element in the set. With this information, the systems seeks to make predictions to the response values for a new dataset. It is common practice to test the performance of the system with a test dataset before applying the model in real-world problems. Usually, a larger training dataset will yield a more robust learning algorithm. Supervised learning algorithms can be divided in two categories, namely classification, where data is assigned to different categories, and regression, for problems with a continuous set of possible values (MathWorks 2015a). Unsupervised learning algorithms, on the other hand, aim to draw inferences without training with previously labelled datasets. The most common method is cluster analysis, which is

[1] More information about this groups can be found at http://research.microsoft.com/en-us/groups/nlp/ and http://research.google.com/pubs/NaturalLanguageProcessing.html, respectively

an exploratory method used to find hidden patterns or groupings in a given dataset. A measure of similarity must be defined and it is usually a distance metric, which can be a fixed one like the Euclidean Norm, or a non-deterministic one, like a probability distribution (MathWorks 2015b). The third strategy, inductive learning, seeks to establish a general rule from a set of observations. This means that given a dataset it will analyse it and produce its own rules for determining whether two elements are 'close' or not. This type of algorithm removes the need to specify a similarity measure with which the data should be compared (Cardie 2015).

From the inductive learning paradigm, and taking from the nerual networks models, *deep learning* has become the new standard in machine learning research. Bengio, Goodfellow and Courville (2015) describes their goal as "to allow computers to learn from experience and understand the world in terms of a hierarchy of concept, with each concept deïňĄned in terms of its relation to simpler concepts," which in turn will "allow the computer to learn complicated concepts by building them out of simpler ones. If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers, " and hence the name. This way of modelling usually involves multi-layered neural networks which make it complicated to thoroughly understand what goes on.

This discussion should grant enough background into machine learning strategies in order to be able to discuss whether this field can in fact be regarded as an extension of our minds in Clark's sense, a discussion that will be presented on the next section.

Learning machines are not extended minds

In order to argue whether machine learning is indeed a form of extended mind, it is worth recalling Andy Clark's rules for determining the possibility of an external element to couple strongly enough with the internal processes of cognition to be considered as an extension of our minds, if only for a specific task. Paraphrasing the four rules, an environmental resource must achieve the following: (1) to be available and usually relied upon; (2) to posses information that is automatically endorsed and not usually criticised, as trustworthy as biological memory; (3) the information must be accessed with ease; and (4) the information must have been previously endorsed *consciously*.

Let us tackle them one by one to reach an adequate conclusion. Firstly, we shall discuss availability. It is a well known fact that at least of the technologies that we currently use

on a daily basis incorporates machine learning as fundamental part of their software design. Not everybody has a mobile phone nearby every day at all times, but just as the fact that not every person needs a notebook to keep track of their memories does not invalidate the case of Otto using it to extend our minds, we can argue that for *those* people who indeed carry their phones with them every moment, this resource is *reliably available and typically invoked*. It is not the mobile phone *per se* that is the object of discussion in this example, but the machine learning algorithms programmed into it or accessible through it which concern us.

Secondly, I will skip to number three on the list because it is very much related to the first one. We can access all kinds of information from our mobile phones local and through the internet. For local applications on the phone, the answer is without second thought yes and for access to remote resources, wireless and mobile phone networks are now reliable enough to not second guess whether we have access to them. Even cases where access is momentarily impossible, when you have no coverage or have run out of battery, Clark could argue that they are comparable to getting drunk on a night out and not having complete access to your internal cognitive capabilities for a given amount of time.

Thirdly, automatic acceptance must be discussed. Imagine that this person, lets call her Anna, who has her mobile phone readily available, every time she needs a piece of information will reach into her pocket for it. This time, she is looking to buy a book but is not sure which she might like so she logs in to her Amazon account and looks through the list of suggestions made by Amazon's recommendation algorithm. This is the crucial point for the argument, if she were to "automatically endorse" the suggestions, she would buy, without further analysis, one of the books in that list and most likely the top one, since it represents the one the algorithm believes is best for her. However, instead of just buying the book Anna, like most people, is likely to go through the comments and reviews section before making up her mind. This example does not meet the second condition. It may seem to harsh to discard all machine learning algorithms with just this one example, but Amazon is widely regarded as one of the most successful cases in this area (Gomes 2015) and if this algorithm cannot pass this test, simpler ones do not stand a chance.

Lastly, for the sake of completeness, I will turn to conscious endorsement of the information. There is no scenario in which this can be applied to machine learning. It

is obvious with unsupervised learning algorithms that this is not the case, since the computer reaches conclusions by itself after only being given a set of rules. The fact that it follows a given rule does not mean that the information produced by the algorithm is accepted as true. If we go back to Anna's example, she never put the books in the recommendation list herself and if she had, no machine learning algorithm would be required to generate the list. We can take another example given by Hayles (2012, p.75-79), where she proposes the "Literature +" approach for reading and text analysis that includes a combination of human and machine learning. The system would work as follows: a machine would analyse massive amounts of text and from there draw some conclusions regarding relationships through strategies mostly of unsupervised learning. After this clusters are pointed out, a person will analyse them and draw conclusions from them. Let us assume that the first three conditions are met in this example so, if the fourth one is true it will be a true instance of the extended mind. But it is impossible for this system to pass this test because the information present was never "consciously endorsed" by the person using it for the analysis of the texts rather it was generated *by the computer itself* in order to be presented to a person. This is the most difficult test to pass for machine learning and so far there is not a system I have come across that could successfully do so.

Conclusion

An historical account of both machine learning and the extended mind paradigm have been presented. This account has provided the tools to better understand both concepts in order to properly answer the question *can machine learning be an instance of the extended mind?*. That anything conunts extended mind concept is not immediately easy to determine as even the definition of mind thus far is shaky at best, amongst philosophers as well as neuroscientists, but the computational explanation is the most widely regarded as true. However, we must be careful and consider the warning by Searle (1984, p.44):

> Because we do not understand the brain very well we are constantly tempted to use the latest technology as a model for trying to understand it. In my childhood we were always assured that the brain was a telephone switchboard... [Charles] Sherrington... thought that the brain worked like a telegraph system. Freud often compared the brain to hydraulic and electro magnetic systems. Leibniz compared it to a mill, and... some of the ancient

> Greeks thought the brain functions like a catapult. At present, obviously, the metaphor is the digital computer.

Nevertheless it is fortunate that Andy Clark provided us with a specific set of conditions, all of which must be met if we are to consider any external tool as an extension of our minds. It was clearly demonstrated that machine learning at this time can only fit two of them and the remaining two are not likely to be tackled in the foreseeable future, especially when current research, according to private conversations with Pedro Mercado —a PhD student in the Machine Learning Group at Saarland University[2]— is targeting inductive and deep learning algorithms, which hope to completely take the human *out* of the machinic though process rather than coupling with it. Not even the highly cooperative "Learning +" scenario proposed by Katherine Hayles in which a machine analyses a text to discover clusters and hidden relationships in it and is afterwards passed on to a person to do further analysis meets the last criterion. An argument could be that this fourth property was not part of the original extended mind proposal and was only added as an answer to attacks on the original thought exercise about Otto's notebook. Nonetheless it was included to provide a better understanding and more rigorous approach to determine whether certain external tools can couple with the internal thought processes in the appropriate way as to enable an extension of the mind, and this condition is not met by current machine learning.

[2]http://www.ml.uni-saarland.de/people/lopez.htm

References

Adams, Frederick and Kenneth Aizawa (2001). 'The bounds of cognition'. In: *Philosophical Psychology* 14, pp. 43–64.

— (2010). 'Defending the Bounds of Cognition'. In: *The Extended Mind*. Life and Mind: Philosophical Issues in Biology and Psychology. MIT Press, pp. 67–80.

Belaval, Yvon (2015). *Gottfried Wilhelm Leibniz*. URL: http://www.britannica.com/EBchecked/topic/335266/Gottfried-Wilhelm-Leibniz (visited on 22/04/2015).

Bengio, Yoshua, Ian J. Goodfellow and Aaron Courville (2015). 'Deep Learning'. Book in preparation for MIT Press. URL: http://www.iro.umontreal.ca/~bengioy/dlbook.

Braddon-Mitchell, David and Frank Jackson (1997). *Philosophy of Mind and Cognition*. Blackwell.

Cardie, Claire (2015). *Inductive Learning*. URL: https://www.cs.cornell.edu/courses/CS4740/2012sp/lectures/ml-basics-ai-lecture-4pp.pdf (visited on 23/04/2015).

Clark, Andy (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.

— (2010). 'Memento's Revenge: The Extended Mind, Extended'. In: *The Extended Mind*. Life and Mind: Philosophical Issues in Biology and Psychology. MIT Press, pp. 43–66.

Clark, Andy and David Chalmers (2010). 'The Extended Mind'. In: *The Extended Mind*. Life and Mind: Philosophical Issues in Biology and Psychology. MIT Press, pp. 27–42.

Clark, Andy and David J. Chalmers (1998). 'The Extended Mind'. In: *Analysis* 58.1, pp. 7–19.

Damasio, A.R. (2004). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Vintage. ISBN: 9780099421832.

Fernandes, Luis (2015). *A Brief History of the Abacus*. URL: http://www.ee.ryerson.ca/~elf/abacus/history.html (visited on 22/04/2015).

Freiberger, Paul A. and Michael R. Swaine (2015). *Pascaline*. URL: http://www.britannica.com/EBchecked/topic/725527/Pascaline (visited on 22/04/2015).

Gomes, Lee (2015). *Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts*. URL: http://spectrum.ieee.org/robotics/artificial-intelligence/machinelearning-maestro-michael-jordan-on-the-delusions-of-big-data-and-other-huge-engineering-efforts (visited on 23/04/2015).

Hayles, N.K. (2012). *How We Think: Digital Media and Contemporary Technogenesis*. University of Chicago Press. ISBN: 9780226321400.

MathWorks (2015a). *Supervised Learning*. URL: http://uk.mathworks.com/discovery/supervised-learning.html (visited on 23/04/2015).

— (2015b). *Unsupervised Learning*. URL: http://uk.mathworks.com/discovery/unsupervised-learning.html (visited on 23/04/2015).

McCarthy, J. et al. (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.

Menary, Richard (2010). 'Introduction: The Extended Mind in Focus'. In: *The Extended Mind*. Life and Mind: Philosophical Issues in Biology and Psychology. MIT Press, pp. 1–26.

Miłkowski, Marcin (2013). *Explaining the Computational Mind*. MIT Press.

Mitchell, Thomas M. (1997). *Machine Learning*. 1st ed. New York, NY, USA: McGraw-Hill, Inc.

Neumann, John von (1993). 'First Draft of a Report on the EDVAC'. In: *IEEE Ann. Hist. Comput.* 15.4, pp. 27–75. ISSN: 1058-6180.

Noë, A. (2009). *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons from the Biology of Consciousness*. Farrar, Straus and Giroux. ISBN: 9780809074655.

Putnam, H. (1981). *Reason, Truth and History*. Philosophical Papers. Cambridge University Press. ISBN: 9780521297769.

Ryle, Gilbert (1973). *The Concept of Mind*. Penguin Books. ISBN: 9780262018869.

Searle, J.R. (1984). *Minds, Brains, and Science*. Reith lectures. Harvard University Press. ISBN: 9780674576339.

Shannon, Claude (1948). 'A Mathematical Theory of Communication'. In: *Bell System Technical Journal* 27, pp. 379–423, 623–656.

Sloman, A. (1978). *The Computer Revolution in Philosophy: Philosophy, Science, and Models of Mind*. Harvester Studies in Cognitive Science. Harvester Press. ISBN: 9780855273897.

The Extended Mind (2010). Life and Mind: Philosophical Issues in Biology and Psychology. MIT Press.

Turing, A. M. (1937). 'On Computable Numbers, with an Application to the Entscheidungsproblem'. In: *Proceedings of the London Mathematical Society* s2-42.1, pp. 230–265. DOI: 10.1112/plms/s2-42.1.230. URL: http://plms.oxfordjournals.org/content/s2-42/1/230.short.

Turing, Alan M. (1950). 'Computing Machinery and Intelligence'. In: *Mind* 59, pp. 433–460.

Weiner, Charles (1973). *Interview with Dr. Richard Feynman*. URL: http://www.aip.org/history/ohilist/5020_5.html (visited on 21/04/2015).

```
File: CTEssay.tex
Encoding: utf8
Words in text: 5792
Words in headers: 25
Words outside text (captions, etc.): 31
Number of headers: 8
Number of floats/tables/figures: 1
Number of math inlines: 1
Number of math displayed: 0
Subcounts:
  text+headers+captions (#headers/#floats/#inlines/#displayed)
  1+6+0 (1/0/0/0) _top_
  297+1+0 (1/0/0/0) Subsection: Introduction
  1065+2+0 (1/0/0/0) Subsection: The Mind
  1029+2+0 (1/0/0/0) Subsection: Extended Mind
  1071+2+7 (1/1/1/0) Subsection: About computers
  1043+5+20 (1/0/0/0) Subsection: Artificial Intelligence and Machine Learning
  865+6+0 (1/0/0/0) Subsection: Learning machines are not extended minds
  421+1+4 (1/0/0/0) Subsection: Conclusion
```