Machine Autonomy in the Social Space

Addressing the Skeptic

The revival of the prospect for artificial intelligence is brewing debate in many fields as the “AI winter” thaws. After a long period of reduced funding slowed the breakthroughs and hit an all-time low around 1990, the topic is back on the radar. This is agreeably attributed to recent and mounting progress in the subfield of machine learning. While many tech enthusiasts hold that machine learning is the golden brick road towards taking AI out of science fictions books and making it reality, skeptics challenge that even the continued success of machine learning may not be enough to declare victory.

In this essay, I will argue that the finer employments of this technology have already begun to disarm the steadfast skeptic, as it has taken form in surprising and complex manners that are missing from the classic debate. My purpose in writing this is to address the stronger skeptic, who maintains a denial of machine autonomy, even with the concessions to certain abilities of these systems. In doing so, I will first enforce machine learning and the use of artificial neural nets as authentic learning, which the stronger skeptic may yet allow, but is still a point of contention for weaker skeptics and either way too central to the overall discussion to omit. To attempt to push this argument forward I will analyze the presence of more complex intelligence bodies in the social space, which if I am successful in doing, I claim will give the skeptic little recourse in denying machine autonomy.

Making these assertions, I will dutifully try to stay clear of anything that could be considered science fiction and rather work with examples that are available and in use today, as it is my hope to claim that these complex systems in their current state have already begun to take the form of the artificial intelligence the skeptic argues can never exist, and the tech enthusiast still believes to lie in fiction. In addressing the strong skeptic, I take Lovelace's objection as the most concise and powerful statement against machine autonomy. I will give proper attention to my interpretation of the objection after covering some ground on machine learning.

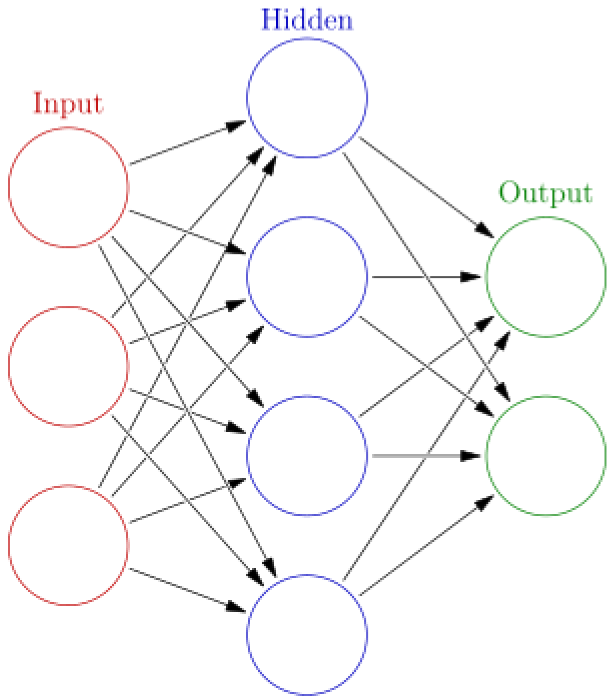
Machine Learning

What constitutes a machine "knowing" something in the first place may have been a point of contention until recent work in computer science. Surely, we do not credit the physical possession of data as being substantial enough to be considered knowing it. We demand that a machine be able to recall and reference data on its own, including knowing where to look for it, similar to our own standards of "knowing" something. This operation is no doubt "thinking." In taking this definition of thinking, we can move forward into more complicated cognitive functions such as learning. More abstract definitions of thinking are not successful in denying this feature of computers; Turing was well-founded in stating that this original question of "Can machines think?" is "too meaningless to deserve discussion.[1]" Furthermore, even Lovelace was willing to grant that machines could be capable of such performances, and thus is not a point of debate for this essay.

In order to explore some of the more complex and trending strategies computer science employs, I think it is logical to start first with exploring the cornerstone of these new models, which is the technique referred to as “deep learning.” Deep learning is inspired by our own fashion of thinking, involving neural activity mainly in the neocortex, the most recently developed area of the brain. “The software learns, in a very real sense, to recognize patterns in digital representations of sounds, images, and other data.[2]” The process involves a series of input neurons being tied (in every possible manner) to another “layer” of neurons before the output. These intermediate layers are referred to as “hidden” because they belong to the proverbial black box of the system and make the system fundamentally opaque.

---

[1] Turing, Alan M. "Computing Machinery And Intelligence." Mind 59.236 (1950): 433-60. Web. 5 May 2017.
[2] Hof, Robert D. "Deep Learning." MIT Technology Review. 29 Mar. 2016. Web. 07 May 2017. <https://www.technologyreview.com/s/513696/deep-learning/>.



Artificial neural network with layer coloring.
en.wikipedia.org/wiki/Artificial_neural_network

Artificial neural networks are decades old, “developed in the 1950s not long after the dawn of AI research” and looked like a promising direction for AI research.[3]” With the onset of deep learning, the number of hidden layers expanded dramatically, and the capacities of these machines grew in turn. The methods of tweaking the networks changed as well. Neurons are “weighted,” randomly at first, but changed throughout the training cycle of the machine. The weights “determine how each simulated neuron responds - with a mathematical output between 0 and 1.[4]” Through techniques such as backpropagation, the network makes adjustments to the weights of its neurons until it creates the desired output[5].

[3] Ibid.
[4] Ibid.
[5] Ibid.

Appreciating some of the more complicated math involved in an ANN can give an even more intimate understanding of their function. One can gain an operational understanding without wading too far into complex math. In backpropagation, each neuron can be imagined as having an error surface (all possible returned values from the neuron) in the shape of a hyperparaboloid. However, in most problems “the solution space is quite irregular with numerous ‘pits’ and ‘hills’ which may cause the network to settle down in a ‘local minimum’ which is not the best overall solution.[6]” Since this modeled error cannot be known *a priori*, analysis of a network requires a large number of runs to determine the best solution. Most networks have have built-in methods for dealing with finding a good instance of the network, which involves modifying the speed in which the error surface is traversed to better avoid local minima, but that is beyond the scope of this paper[7]. However, this does give us the ability to appreciate the uniqueness of these neural nets, and acknowledge, at least at the explicit level, the designers limits in programming one.

As stated before, this technique has been successful in recognizing a multitude of data. Perhaps more surprising is that running one of these neural networks in reverse will create something that looks like a viable input. For example, if we take an artificial neural net designed to translate handwritten single integers into a given arabic numeral (0-9) and run it in reverse, it will effectively “draw” a number. This is a rather surprising

[6] "A Basic Introduction To Neural Networks." University of Wisconsin-Madison. Web. 07 May 2017. <http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html>.
[7] Ibid.

unintended consequence of the network. It suggests that the ANN hasn't simply created a lucky estimate for lighting up the right corresponding box and satiating us with having the right answer, rather, the network itself actually has some impression of what it believe, say "4" looks like.

This is shockingly similar to human behavior. In training a child to draw numbers or letters correctly, they may reference an alphabet strung across the top of a classroom blackboard. Through trial and error, a child is typically able to generate a symbol acceptably similar to the one offered. However, they can never create exactly the image above the board. This is never a point of contention in deciding whether or not a child has successfully reproduced the number 4. Even if a human has a so-called "perfect" (geometrically speaking) symbol in their head, they may not be able to render it. What is important is that we are so acceptably similar to the ideal character that it is useful in communication. The same then, applies to machines.

This does little to dissuade the strong skeptic, mentioned earlier, as Lovelace herself would concede that such performances are possible by machines. And furthermore, that even though a programmer (I refer to the designer of these machines as a programmer only because such is the role of the proverbial "designer" in modern examples) may not be able to exclusively program a neural network to identify the number 4, in the network having that goal as prescribed by the designer, it has some arcane residue coating it.

In my interpretation of Lovelace, that residue is the persistence of the original “order” in the performances of the machine. And perhaps it is only a convenient side effect that an ANN can be operated in reverse to “draw” a symbol. The performances of these networks is not the matter of concern here, as they are already conceded by Lovelace. An analogous performance to any trivial ability, regardless of computational cognition required, is not enough to dismount the objection. To do so, we will have to take the understanding of machine learning given here and go some level of abstraction higher.

Machines in the Social Space

To begin here, I believe I must first introduce what I am considering the social space. The idea of a social space is not unique to humans, and indeed, even precedes them as other species such as baboons are social. In the same way as John Wise did, I refer to our human social space as that which “consists of both that which subjects directly manipulate physically (technology) and that which they manipulate incorporeally (language).[8]” Additionally, actor-network theory charges that technology is always a social actor. “Both animate and inanimate actors bend space around them (Callon & Latour, 1981, p. 286).[9]” I aim to explore that in being both the technology, and in interacting with us incorporeally via language, we can make a strong argument against Lovelace’s objection and the strong skeptic.

[8] Wise, John Macgregor. Exploring Technology and Social Space. London: Sage Publications, 1997. Print.
[9] Ibid.

I will accept the strict social model stated as our definition of the social space, as it is not the goal of this essay to debate the social sciences. Furthermore, if I am successful in making a case for the undeniable presence of these machines in the social space, other definitions of such space should not detract from the greater point. In exploring our new interactions with machines, it is first important to give some attention to their use of language. Machine use of language has recently had tremendous success as a result of deep learning and similar computer and cognitive science breakthroughs. These successes are known commonly under the label of natural language processing.

Natural language processing is generally the practice of using neural conversational models to synthesize speech. Much like in the fashion of the ANNs explored earlier, these are “purely data-driven systems trained end-to-end on dialogue corpora.[10]” While typically successful contextually, such systems tend to generate short, generic responses. This model is known as the sequence-to-sequence model. While its applications are widely employed, that are still notable challenges. In seeking more conversational speech, much like the kind envisioned necessary for passing the Turing test, one may encourage these models to give longer responses. “[W]hen longer responses are explicitly encouraged (e.g. via length normalization),
they tend to be incoherent (“The sun is in the center of the sun.”), redundant (“i like cake and

[10] Shao, Louis, and Stephan Gouws. "Generating Long and Diverse Responses with Neural Conversational Models." Google Research (2017): n. pag. Web. 7 May 2017. <https://arxiv.org/pdf/1701.03185.pdf>.

cake"), or contradictory ("I don't own a gun, but I do own a gun.").[11]"

Fortunately, researchers at Google have successfully generated much longer and diverse responses using a new model they call the glimpse-model.[12] Google has been integrating their advances in natural language processing across a wide variety of technologies that are in constant use. The result is some highly beneficial data processing in combination with being able to convey it to us via natural language. This phenomenon takes quite a few forms. Google search itself has become increasingly powerful and has seen many new features come as a result, such as the knowledge graph. The knowledge graph is a complex web of the information that exists online. It can make contextual connections, so that in searching "famous jazz musicians," Google returns its understanding of the most influential people. This is done in a fashion similar to how we may list them, in that there is no explicit correct list and the order is not of import. What is important though, is that the knowledge graph itself understands the context and can present the results to us in natural language.

While the knowledge graph itself does not necessarily showcase natural language processing itself, when paired with other services the use becomes more apparent. The Google Assistant, the company's take on the much-buzzed-about virtual artificial intelligence agent, frequently makes use of such technology. The assistant, independent of the knowledge graph, can reference it and supply natural language

---

[11] Ibid.
[12] Ibid.

responses to the user from a voice search. There are quite a few important things happening here. First, the agent is listening and interpreting an audible query supplied by a human. Second, it is able to construct a context that this query exists in, and search it. And finally, it is able to supply a genuine, contextually-correct response in natural language. Perhaps more impressive, though, is it's ability to keep the context. After asking the assistant "Who are some famous jazz musicians?" and getting a response, you may want to respond with "What about rock?" The assistant retains the context of the conversation and can offer a new response.

Every one of these "searches" is a query-response linguistic interaction between humans and computers. We conduct these conversations en masse on a daily bases, and Google is conducting these interactions largely on its own. That is not to suggest that search passes the Turing test, as it surely does not, but these interactions are undoubtedly the incorporeal linguistic exchanges I mentioned earlier, and clearly makes a valid case for its presence in the social space. Furthermore, in the context of Lovelace's objection, I do not think it can be said that the "order" here is easily defined. This complex network of intelligence is making it's own contextual understanding of the world and figuring out its own approach to the social space, constantly.

Having this relationship with us in a manner in which machines are naturally integrated in the social space, is, I think, a strong strike at Lovelace's objection. Where she states, regardless of the function of a machine (In other words, conceding that one may be able to replicate any complex process we theorize in a machine) the machine is only able to do what we have ordered it to do. The complex intelligence that I am examining, combined with its undeniable presence in the social space transcends any such one of these “orders” that Lovelace’s objection could have possibly envisioned. That is to say, it is often stated in favor of Lovelace’s objection the distillation of a machine’s operations (regardless of its matter to impress) into a simple goal (or “order,” as originally stated by Lovelace). I claim, that when the “order” is so abstract that it begins to make postulations at the same level of abstraction that we use when evaluating our own autonomy and agency, Lovelace’s objection breaks down. As these “orders” of these machines are no longer simple goals in the classic image of a designer and his system any longer. The goals or “orders” of these complex intelligence bodies in the social space are a direct order of the social space and society itself. In this manner, I feel that Lovelace’s objection either fails to refute the intelligence of these machines, or perhaps worse, states that we are thereby bound by a similar “objection” that would apply to our own intelligence.

Works Cited

- Turing, Alan M. "Computing Machinery And Intelligence." Mind 59.236 (1950): 433-60. Web. 5 May 2017.
- Hof, Robert D. "Deep Learning." MIT Technology Review. 29 Mar. 2016. Web. 07 May 2017. <https://www.technologyreview.com/s/513696/deep-learning/>.
- "A Basic Introduction To Neural Networks." University of Wisconsin-Madison. Web. 07 May 2017. <http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html>.
- Wise, John Macgregor. Exploring Technology and Social Space. London: Sage Publications, 1997. Print.
- Shao, Louis, and Stephan Gouws. "Generating Long and Diverse Responses with Neural Conversational Models." Google Research (2017): n. pag. Web. 7 May 2017. <https://arxiv.org/pdf/1701.03185.pdf>.
- Estrada, Daniel. "Rethinking Machines: Artificial Intelligence Beyond the Philosophy of Mind." Thesis. University of Illinois, 2014.