Will Machines Ever Think Like Humans?

by Jeff Riley

Editor's Introduction

What is "human intelligence?" What is thinking? What does it mean to "think like a human?" Is it possible for machines to display human intelligence, to think like humans? This article explores these questions, and gives a brief overview of some important features of the human brain, and how computer scientists are trying to simulate those features and their ability to "think." The article answers some questions, but asks more—finishing with questions for readers to consider.

Will Machines Ever Think Like Humans?

by Jeff Riley

What does it mean to "think"?

Alan Turing, mathematician, computer scientist, and philosopher, began his 1950 paper "Computing Machinery and Intelligence" [1] with the sentence: "I propose to consider the question, 'Can machines think?'"

Recognizing that the term "think" may be difficult to define, Turing instead proposed the "imitation game" as a means to determine: "…whether there are imaginable computers which would do well (at the imitation game)" [1].

Turing considered the imitation game, and this new question, a proxy for the question "Can machines think?" The imitation game later became the "Turing test," a test of a machine's ability to exhibit intelligent behaviour indistinguishable from that of a human. The details of the imitation game are not important here—I refer interested readers to Turing's 1950 paper. The salient point is that Turing considered a machine that displayed intelligent behavior indistinguishable from that of a human to be the same as, or a reasonable proxy for, the machine "thinking."

The philosopher John Searle, in his 1980 paper "Mind, Brains, and Programs," proposed the "Chinese room" thought experiment in order to challenge the notion that "strong artificial intelligence (AI)" could enable machines to think [2]. As for the imitation game, the details of the Chinese room are not important here—interested readers should refer to Searle's paper.

The point Searle was making with the Chinese room is that, in his estimation, however well a machine is constructed/programmed, it doesn't really understand anything—it can only simulate knowledge, and simulating knowledge is different from thinking. In his article, Searle asserts: "…strong AI has little to tell us about thinking, since it is not about machines but about programs, and no program by itself is sufficient for thinking." He also states: "Any attempt literally to create intentionality artificially (strong AI) could not succeed just by designing programs but would have to duplicate the causal powers of the human brain."

"Intentionality" is a term used by philosophers to describe mental states (thought, belief, hope, rage, desire, etc.) that are "directed" at objects or ideas. In Searle's words: "Intentionality is

best thought of as 'mental representation'… Intentionality is the capacity the mind has to represent objects and states of affairs… Contents of thoughts are intentional states."[1]

To Searle, "strong AI" was AI that could create intentionality—effectively AI that thinks and, necessarily according to Searle, understands. My interpretation of this is that Searle's definition of thinking embodies the mental processes that manifest intentional states.

A number of criticisms of Searle's paper were printed immediately following the paper in *Behavioral and Brain Sciences*, under the banner "Open Peer Commentary" [3]. I think the most illuminating, perhaps because it accords with my own beliefs and criticism of Searle's paper, is the commentary by Douglas Hofstadter, cognitive scientist, physicist, and Pulitzer Prize-winning author, whose research includes concepts such as the sense of self, and consciousness. I encourage those interested to read Hofstadter's criticism in its entirety.

Hofstadter begins his reply with: "This religious diatribe against AI, masquerading as a serious scientific argument, is one of the wrongest, most infuriating articles I have ever read in my life."

He goes on to say: "Searle's trouble is … he has deep difficulty in seeing how mind, soul, 'I,' can come out of brain, cells, atoms."

I think Hofstadter understands, but Searle missed, an important point: Intelligence, or in Searle's terms, intentionality, isn't in the structure of the brain alone—it is generally accepted to be an emergent property of the interaction of the signals propagated through the networks of neurons in the human brain, and the cells that intercept and manipulate those signals. Thought, intentionality, and intelligence—indeed "mind, soul, 'I'"—are generally accepted to be in the "connectedness" of the brain's neurons—in the signal rates and strengths, the excitation levels, and firing thresholds of the neurons.

What Hofstadter refers to as "mind, soul, 'I'" and the related concepts of consciousness, self-awareness, etc. are beyond the scope of this article. I will restrict discussion here primarily to issues surrounding machines displaying intelligent behavior.

On the question of "What does it mean to 'think?'" I will follow Turing. Displaying intelligent behavior indistinguishable from that of a human is a reasonable proxy for (human) thinking, and, in keeping with Turing, that changes the question asked at the beginning of this article from "Will machines ever think like humans?" to "Will machines ever display human intelligence?"

---

[1] What is Intentionality? | Closer to Truth

HUMAN INTELLIGENCE

Turing did not formally define what he meant by "intelligent behaviour indistinguishable from that of a human," and that deficiency has led to some debate and criticism. Here I will try to define what I mean by "human intelligence." Human intelligence is just intelligence as it is displayed by humans—but what is intelligence? There seems to be no single, generally agreed upon definition of intelligence. Robert J. Sternberg, prominent human intelligence researcher and developer of the Triarchic Theory of Intelligence [4], is quoted as saying: "Viewed narrowly, there seem to be almost as many definitions of intelligence as there were experts asked to define it" [5].

Sternberg defines human intelligence as: "[The] mental quality that consists of the abilities to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment" [6].

What is it that sets human intelligence apart from (current) machine intelligence? I think it is the ability to create, to conceptualize, and to contextualize; the ability to recognize a problem or a question and to devise a solution or an answer; it is the capacity for abstract thought; the ability to reflect on ideas and concepts and think beyond the concrete present. But where does this ability come from? What about the human brain is so different from the various AI engines that attempt to emulate the functionality of the human brain?

The Human Brain
The feature of the human brain most obvious to us is its physical form, particularly its structure. Following is a brief description, at a fairly high level, of the structure and features of the human brain.

The human brain is a very complex organ, comprising several structures that work together to produce the functionality we associate with "human intelligence." The three main structures of the brain are the cerebrum, the cerebellum, and the brainstem.

The brainstem, as well as connecting the cerebrum and the cerebellum to the spinal cord (the part of the central nervous system that is not the brain), controls or orchestrates a number of autonomic functions and states: heart rate, breathing, body temperature, wake (and sleep) cycles, swallowing, digestion, sneezing, coughing, and vomiting. The brainstem contains about 1 billion of the brain's approximately 86–127 billion neurons.[2] AI systems typically aren't concerned with simulating autonomic functions, so brainstem functionality is not usually the subject of AI systems.

---

[2] Estimates of neurons vary. See Herculano-Houzel S. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience* 3, 31 (2009); doi: 10.3389/neuro.09.031.2009.

The cerebellum orchestrates voluntary movements (it controls and coordinates the timing and force of voluntary muscle movements), and maintains posture and balance. It is also involved in motor learning, adapting, and fine-tuning motor programs through trial-and-error (e.g. as children learn various motor skills). The cerebellum is thought to be involved in some cognitive functions—such as language, attention, and emotional responses to fear and pleasure—though the involvement of the cerebellum in these functions is not well understood. Most of the volume of the cerebellum is taken up by the cerebellar cortex, and although much smaller than the cerebrum (in human brains) with about 70–101 billion neurons,[3] the cerebellum contains the vast majority of the brain's neurons.

The cerebrum is responsible for higher functions such as abstract thought, reasoning, language and speech, sensory processing (vision, hearing, taste, smell, and touch), emotions, learning, and fine control of movement. The outer layer of the cerebrum is the cerebral cortex, containing some 16–26 billion neurons.[4]

Figure 1. The human brain.


---
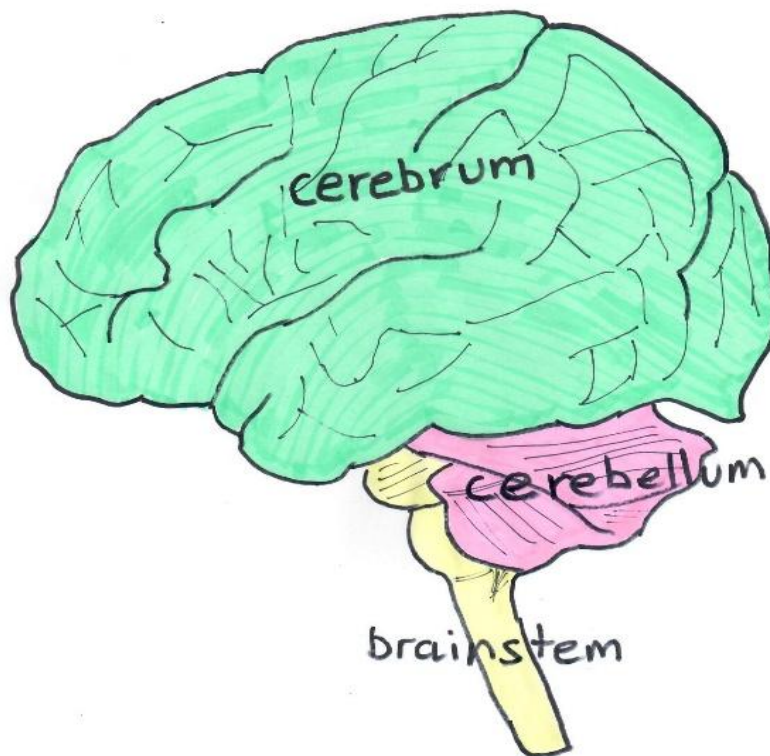
[3] ibid.
[4] ibid.

The cerebral cortex expands, and subsequently folds, as our brains develop in the womb, resulting in the familiar wrinkled appearance of the human brain. The folding of the cerebral cortex allows a larger cortex to fit inside the human skull than would otherwise be the case, and a larger cortex means more neurons, which can, but doesn't always, mean a more advanced brain and increased cognitive abilities. Elephants, for example, have a physically bigger brain than humans, and the cortex of the elephant brain is folded in a similar way to the human brain, resulting in the average elephant brain containing about three times the number of neurons than the average human brain. Elephants are intelligent animals, but by most measures human intelligence is superior to elephant intelligence.

While the cortex of some animal brains is folded (e.g. humans and elephants), the cortex of most animals' brains is not—folding tends to occur in animals with larger brains. The growth of the cortex and the limiting volume of the skull are not the only factors that determine whether folds form in the cortex, the physical properties of the cortex can also make a difference, with thinner regions folding more easily, and so more often. The physical properties and patterns of folding in different regions of the cortex are thought to be linked to the function of those regions. These patterns of folding are consistent across individuals, and even some species, leading to speculation that folding has some underlying function or meaning not yet evident.

The Biological Neuron
The biological neuron is the basic "processing element" of the human brain. The role of the neuron, in very simple terms, is to collect and collate input from other cells (e.g. other neurons or sensory receptor cells), and, if so indicated by the state of the neuron and inputs, transmit information to other cells (e.g. other neurons, muscle, or gland cells).

Figure 2. Biological neuron (pink) showing connection to another cell (yellow).
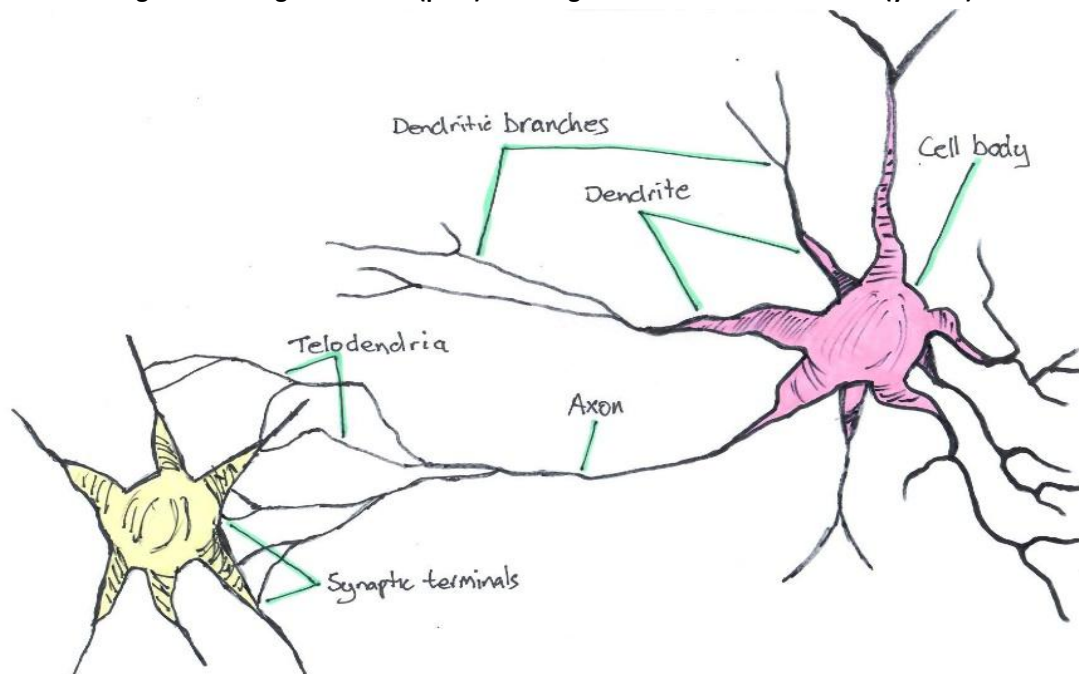
Neurons are connected to other cells via axons and dendrites, specialized projections of neurons that (respectively) deliver and collect information (signals). A neuron communicates with other cells by sending very fast electrical signals, known as "action potentials" on the axon. These signals, the fastest form of intracellular electrical signal in biology, are either transmitted directly to receiving cells (electrical synapses) or stimulate neurotransmitter release at the axon terminal, allowing the signal to propagate to other cells (chemical synapse).

Neurons gather signals from a number of other cells, perform an analysis of those inputs, and pass a signal on to other cells if necessary. The analysis of the input signals performed by the neuron is not well understood, but it is known to vary depending upon the nature of the inputs. In fact, researchers have recently discovered a neuron's dendrites not only gather the input signals, they functionally work together in a way that is adjusted to the complexity of the input [7]. Theories of the operating principles of biological neurons differ. One of the simpler, high-level models has the neuron performing a weighted sum of its inputs, and sending a signal out on its own axon (known as "firing") only if the sum exceeds the neuron's "firing threshold" (and remaining quiescent and propagating no signal if not).

Neurons fire in discrete pulses. Whenever the electrical potential inside the cell body reaches a certain threshold, a pulse is transmitted along the axon. These pulses can be translated into a continuous value—the pulse rate directly affects the rate at which cells connected to the axon receive signal ions. The faster a biological neuron fires, the faster connected neurons accumulate (or lose for inhibitory connections) electrical potential.

Neurons are arranged in a complex series of interwoven networks of interconnected, functionally related neurons. These networks of neurons—the biological neural networks—are the predominant functional features of the human brain, and are thought to be the seat of human intelligence.

How Does the Brain Work?
While we know a little about the structure of the human brain, how it actually works remains largely a mystery. We've discussed what "thinking" is, but how do human brains think? What is thought? How does human memory work? If we want to create systems that mimic (some of) the operation of the human brain, we first need to understand, at least to some extent, how the human brain works.

Neural networks, whether biological or artificial, are effectively "black boxes," in the sense that their operation can be viewed in terms of their inputs and outputs, without any real understanding of how they work internally. Neural networks are not "explainable." In almost all cases it is not possible to extract rules from the network features to explain how the network determines its output from the inputs.

We can experiment to try to get a better understanding of how the human brain works. We can elicit responses from the human brain by poking and prodding it with electricity, chemicals, and

physical probes, and by doing so we can learn a little about the brain's electrochemical signaling and reaction to stimuli. What we learn from such experimentation provides hints about how the brain works, and we can generalize from those hints, but that's a very long way from providing any real, especially detailed, understanding of how the brain actually works—how it engages in abstract thought and other higher-level thinking. Even if we were able to map the responses of every neuron, and associated connection strengths and signaling rates, to a broad range of stimuli, we would still only have an (incomplete) instantaneous picture of a single example of a human brain. Humans are different, and each of us learns, remembers, and forgets. Our brains are dynamic, constantly changing as we adapt and learn. We would need to experiment with many human brains to discover generalized functionality that we could then use to help us develop systems that could precisely mimic the functionality of the human brain. Discovering the underlying operations of the human brain and being able to replicate them in detail is far from being within our grasp in the foreseeable future, but replicating in detail and simulating black-box functionality are very different things—we may only need to simulate black-box functionality in order for machines to achieve human intelligence.

Quantum Processes in the Human Brain
As described above, the conventional view is that intelligence, thought, consciousness, etc. are emergent properties resulting from the interaction of neurons and connections between neurons (and the properties of the neurons and connections, such as signal strength, signaling rate, firing thresholds, etc.). Other ideas have been proposed that challenge the conventional view, and I'll briefly discuss one of those, especially as it relates to the structure and functionality of the human brain.

In a 1987 book, *Ultimate Computing*, anaesthesiologist Stuart Hameroff suggested consciousness originates from quantum states in neural microtubules present inside neurons in the human brain [8]. Roger Penrose—mathematical physicist, mathematician, philosopher of science, and Nobel Laureate in physics—argued in his 1989 book, *The Emperor's New Mind*, human consciousness cannot be represented by an algorithm, and so is not capable of being modeled by a conventional Turing machine (including current digital computers) [9]. Furthermore, Penrose hypothesized that quantum processes in the human brain lead to human consciousness. Penrose and Hameroff later collaborated to develop a biological theory of mind known as Orchestrated objective reduction (Orch OR), postulating that consciousness originates at the quantum level inside neurons in the human brain [10]. The Orch OR hypothesis has generated some criticism and its validity remains controversial.

While Orch OR hypothesizes about the origin of human consciousness, and not about other cognitive processes (e.g. thinking, problem-solving, etc.), it is not directly relevant to this discussion about human intelligence. It also hypothesizes, and invites hypotheses, about some specific structures and functionality of the human brain.

Bandyopadhyay et al. speculated in 2014 that neurons could communicate "wirelessly"—that microtubule-based quantum coherence could extend between neurons that are not physically

connected [11]. Wireless communication between neurons is certainly an intriguing prospect, but at this point it is still speculation. Hameroff and Penrose favor their original "gap junction" proposal, which requires communicating neurons be in physical contact, over the wireless transmission mode proposed by Bandyopadhyay et al., expressing doubt that the hypothesised wireless transmission would be capable of transmitting the required superimposed quantum-states [12].

SIMULATING HUMAN INTELLIGENCE

Much of the work in the AI community related to simulating human intelligence has focused, and continues to focus, on artificial neural networks (ANNs)—mathematical models that are inspired by the structure of animal brains and, in some ways, the functionality of the human brain. The reason for the focus on ANNs is fairly clear: Evolution has already solved the problem of developing human intelligence. Most of the human brain is composed of many biological neural networks which, as stated above, are thought to be the seat of human intelligence.

We should note, however, while evolution has found a solution that manifests human intelligence, it may not be the only solution, or even the best solution—we only need to look at the placement of the optic nerve at the back of the eye to attest to that. Evolution doesn't employ an exhaustive search—it stops looking after it has found a solution that conveys an evolutionary advantage over competitors—so any solution it finds may not be the best, nor the most efficient. (Both are somewhat subjective terms; what makes one solution better than another, and how is efficiency measured?). Moreover, evolution is a slow, incremental process, and can only use the tools and building blocks at its disposal, with any modification to those tools and building blocks evolving over a very long time. Having found a solution that works—biological neurons and neural networks—evolution has spent millions of years incrementally refining that solution. Humans, on the other hand, are able to rapidly (in comparison to evolutionary timeframes) develop new tools and building blocks.

ANNs are just one of the tools developed by humans, and while the result of evolution is the only exemplar we have, there is no *a priori* reason to suggest replicating the evolutionary solution is the best method of developing computers that can display human intelligence Is it possible that AI with a different architecture from that of the human brain (that is, not an ANN) could display human intelligence? Or do we believe human intelligence is an artifact of the architecture of the human brain, and can only be replicated by replicating that architecture? Is the capacity for human intelligence confined to brains—natural or artificial—that look like human brains, or is it possible that other techniques, or combinations of techniques (e.g. decision trees, expert systems, etc.), might simulate human intelligence at least as well as ANNs? Answering these questions could fill an entire book, and I'll leave them for another time. For now I'll focus on ANNs; there are good reasons for doing so [13]:

9

- Expert systems that use symbolic representations usually become slower with a larger knowledge base, because larger sets of rules need to be traversed. Human experts, however, usually become faster. Maybe a non-symbolic representation (as it is used in natural neural networks) is more efficient.

- Despite the fairly long switching time of natural neurons (in the order of several milliseconds) essential cognitive tasks (like recognizing an object) are solved in a fraction of a second. If neural processing were sequential, only about 100 switching operations could be performed ("100-step rule"). Hence, high parallelization must be present, which is easy to achieve with neural networks, but much more difficult to implement with other approaches.

Artificial Neural Networks

As described above, an artificial neural network is a mathematical model inspired by the structure of animal brains and, in some ways, the functionality of the human brain. As discussed above, animal brains, most particularly human brains, are composed of a number of biological neural networks, and those biological neural networks themselves are composed of large numbers of biological neurons. ANNs are constructed (most often in software simulations, though hardware ANNs do exist), using artificial neurons that model, to varying degrees, the behavior of biological neurons.

The artificial neuron. The fundamental building block of ANNs is a crude analogue of a biological neuron, called the artificial neuron.

Based on the Threshold Logic Unit (TLU) proposed by McCulloch and Pitts, an artificial neuron takes one or more inputs, each either excitatory or inhibitory, and tests a function of those inputs against a threshold value—if the threshold is exceeded, the artificial neuron "fires," producing an output signal [14].

Input(s) to the artificial neuron are analogous to the signals presented at the dendrites and synaptic terminals of the biological neuron, and are typically the product of the input signal strength and the input weight (lower left box of Figure 3). The sum of the input signals (for simplicity we ignore the bias input often seen in artificial neurons—refer to the literature for details) is passed through an "activation function" (sometimes "transfer function") to determine the artificial neuron output. Typical outputs of an artificial neurons are 0 (or -1) or 1 for neurons that use step functions as their activation function (upper right box of Figure 3), or a floating-point value between 0 (or -1) and 1 for neurons that use sigmoid-type functions as their activation function (lower right box of Figure 3), but other types of activation functions are also used. An artificial neuron is a linear classifier; it predicts whether or not the data presented at its inputs belong to a particular class based on a linear combination of the input values and weights. The output of an artificial neuron is analogous to the signal propagated on the axon of the biological neuron.

Figure 3. Artificial neuron.
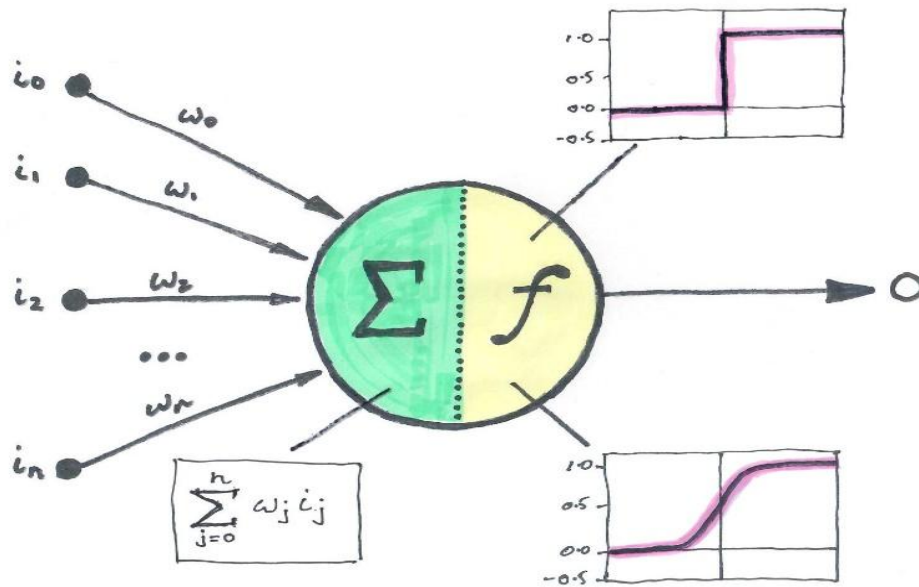
The perceptron. On its own the artificial neuron cannot learn: The activation function and input weights are predetermined and fixed. Rosenblatt's perceptron algorithm describes a method of iteratively adjusting the input weights of an artificial neuron by presenting pre-classified data ("training examples") at its inputs and comparing the output to the known classification [15]. If the input data is linearly separable, and enough training examples are available, the perceptron algorithm guarantees that a set of input weights can be found that will cause the artificial neuron to successfully predict the classification of unseen data (i.e. data that is not part of the training examples used to "train" the input weights). In the context of ANNs, a perceptron is an ANN with a single artificial neuron that uses the unit step function as its activation function (the unit step function results in values of 0 or 1).

In 1969 Marvin Minsky, cognitive and computer scientist, and co-founder of the Massachusetts Institute of Technology's (MIT) AI laboratory, and Seymour Papert, mathematician, computer scientist, and MIT researcher, published their book *Perceptrons* [16]. The book discussed some of the strengths of the perceptron, but also highlighted some major limitations. It is generally thought that Minksy and Papert's criticism of the perceptron and pessimistic predictions in the book contributed to the so-called "AI winter" of the 1980s. After publication of the book, the direction of AI research changed, with focus shifting from the perceptron and ANNs to "symbolic" AI systems such as expert systems. Later, the advent of the "Hopfield net" [17] and, separately, the multi-layer network training algorithm "backpropagation" [18] helped revive research into ANNs. Universal approximation theorems have since shown that ANNs are capable of representing a wide variety of functions when given appropriate connection weights. While performance ANNs can be sensitive to small changes in input patterns, careful selection of network topology and training data goes a long way to alleviating those problems and allowing ANNs to generalize well.

In 2009, neuroscientist Henry Markram, claimed he was going to simulate the human brain in a computer. In 2013, the Human Brain Project, with Markram as one of the original executive committee members, was started with funding of 1 billion euros. The project is a long-term research project to allow researchers to advance knowledge in the fields of neuroscience, computing, and brain-related medicine. More than 10 years after the start of the project we don't yet have a human brain simulated in a computer, leading to some criticism of the project. The failure of the project to deliver on the 2009 claim by Markram speaks more to the need for realistic expectations than it does to the efficacy of ANNs.

ANNs are now a major focus of AI research, particularly in the reasonably nascent field of "deep learning."

The network. An artificial neural network is an interconnected group of artificial neurons (see Figure 4). When used in a network, an artificial neuron is usually referred to as a "processing element," or more commonly simply as a "node." The nodes in an ANN are most commonly arranged in sub-groups referred to as "layers."

An important, but often overlooked, point to remember is that it is not always appropriate to think of a node in an ANN as being in a one-to-one relationship with actual biological neurons in a biological neural network. Often it is more realistic to consider a single node in an ANN as modeling the behavior of a group of biological neurons in a biological neural network.

While current ANN architectures aim to simulate networks in which the neurons are physically connected (via axons and dendrites), there is no reason they couldn't simulate wireless communications between neurons, even as a black-box process. But since the concept is still very much in the realm of speculation not enough is known about how it might be modeled. Realistically though, one would guess that the method of communication is not as significant or important as the content or parameters of the communication—signal strengths, signal rates, and timings could all be simulated via physical connections in the ANN mimicking wireless connections. I think it is unlikely, though not impossible, that the actual transmission mode (wireless vs wired) is a significant factor on its own.

How is the Human Brain Different from an ANN?

The artificial neuron described above is a simple, crude analogue of a biological neuron. While implementations of ANNs and artificial neurons vary in many ways, including complexity, the operation of the biological neuron is far more sophisticated than that of an artificial neuron. The connections between neurons in a biological neural network, and the signals propagated across those connections, are far more nuanced than in an ANN. Biological neurons have different types of synapses that operate in different ways and at different speeds, the signals passed between neurons can vary in both strength and timing. Biological neurons can, for example, differentiate input signals based on the rate of arrival of the signals, whereas (most) artificial neurons detect and act on the strength of the signal only. The human brain is different because of its sophistication and nuanced processes.

The average human brain is composed of about 86 billion neurons that are interconnected via about 100 trillion connections. Each neuron is, in simple terms, a tiny processor capable of some limited computation. The number of neurons is an important indicator of brain capacity, but there is good evidence that connection strengths—how strongly neurons influence neurons to which they are connected—are the real information stores of the brain. Google's cat-recognizing ANN from 2012 had one billion connections—10 times more than any previous ANN, but 100,000 times fewer than the average human brain. Today supercomputers are struggling with ANNs an order of magnitude or so bigger than Google's 2012 ANN (our state-of-the-art is still more than four orders of magnitude smaller than the human brain in terms of connections). The human brain is different because of its sheer scale.

Most ANNs are constructed with a layered architecture (see Figure 4). While the cerebral cortex is layered, the nature of the layering and the topology of the networks within the cortex layers is very different from that of ANNs. The layered architecture of ANNs is a fixed, rigid structure, where neurons belong to a particular layer, and, in general, all neurons in a given layer "fire" at the same time—so signals move forward (in a feed-forward ANN) layer-by-layer.

Figure 4. Artificial Neural Network (ANN) (weights and biases not shown).
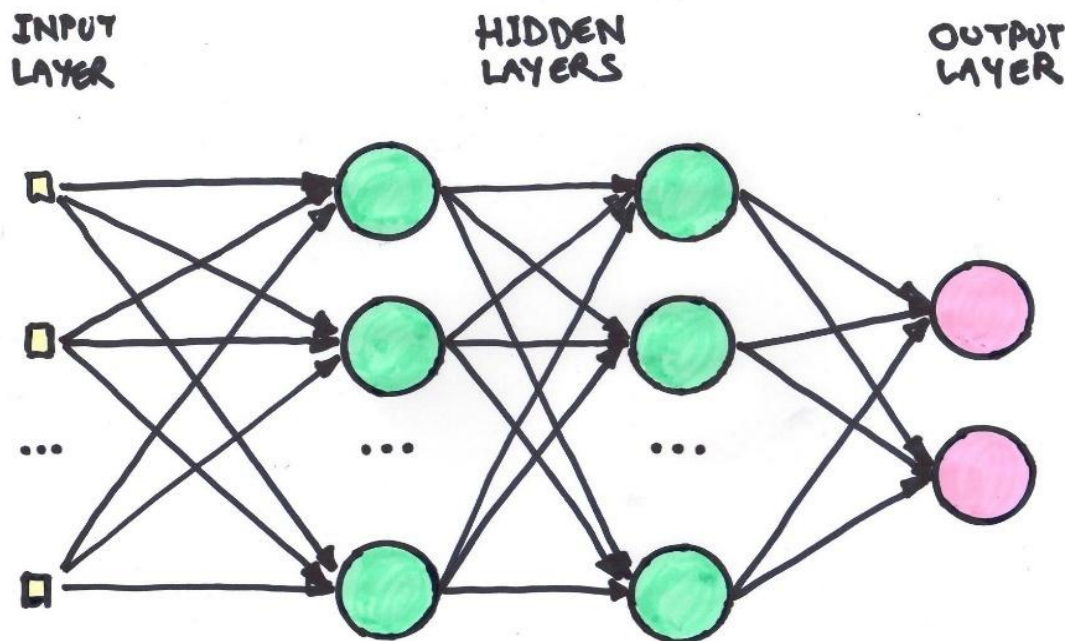
This layered architecture developed at a time when the actual operation of the human brain was less well understood than it is today. More recent studies indicate that neuronal functionality in the cerebral cortex can span multiple layers, allowing neural networks in the cortex to be much more complex and nuanced than previously thought (e.g. Larkum et al [19]). The human brain is different because the topology of the biological neural networks in the

cerebral cortex allows for much greater complexity, diversity, and nuance in the connections between neurons and the distribution of signals throughout the network.

The average human brain has more than 5,000 times the number of neurons than the world's largest ANN with 16 million neurons. We can simulate neurons in software, and by doing that we can scale to very large sizes, but unless the code to fire a neuron and propagate signals over connections runs on a separate processor (or core) for each neuron and connection, firing of neurons and propagation of signals through the network is done largely in a serial fashion—and that takes time. We can mitigate that by splitting the code across many processors running in parallel. Our largest supercomputer has in the order of 10 million processors, most have fewer than 1 million. We call these computers "massively parallel." In the human brain, neurons fire and signals are propagated simultaneously—the human brain has the capacity to have 86 billion processors running concurrently with 100 trillion propagating signals. The human brain is different because it is breathtakingly parallel.

Current supercomputers require somewhere between 2 million and 18 million watts of power while they are running, depending upon the supercomputer and configuration, and fill large data centers. The human brain, on the other hand, requires about 20 watts of power to run (less than an average household incandescent light globe), and fits in the average human skull. Moreover, the human brain is atop a mobile enclosure that provides nuanced inputs from sensors that are able to pre-process input data before forwarding them to the brain for further processing. The human brain is different because it is small, power-efficient, mobile, and is context-aware with respect to its environment.

Embodiment and Environment
ANNs are typically implemented in software and are trained and executed in the context of a computing environment. Some ANNs are implemented in hardware and are deployed into a working environment, but the network is usually trained offline, perhaps in a computer-simulation of the intended working environment, and it is the trained network, implemented in hardware, that is deployed into the real working environment.

Working human brains, on the other hand, are physical, and are always present in the human body, intimately connected to the body and, almost always, all its sensory capabilities. It is in the context of that embodiment, and connection to the body and sensory inputs, that the neural networks in the human brain learn (i.e. are trained) and execute. There are, of course, issues with people who are, for example, without sight and/or hearing, perhaps from birth, or were born without, or have lost, limbs, or who are paralyzed to varying degrees, and so may lack some sensory inputs—but they are not the majority case, and for now I'll leave those issues aside. Human brains are always immersed in their environment and (generally) receiving sensory inputs and, as a result, are always updating their training (creating new connections between neurons, updating connection characteristics, etc.) in their real working environment.

There are several theories surrounding the embodiment of the human brain. The "embodied cognition" theory posits that many features of cognition are shaped by aspects of the entire body—other theories go further and suggest that the "mind" doesn't reside solely in the brain, or even the body, but extends into the physical world (the "extended mind" thesis).

I think the embodied cognition theory is close to the reality of how the human brain uses its knowledge of its environment, through the body's interaction with the environment and translation of those interactions into sensory inputs to the brain, to construct a representation of its "world," and gain knowledge and understanding from that representation.

An ANN can be deployed into a physical robot that has myriad hardware sensors that are able to provide sensory input to the ANN. Robots can sense electromagnetic radiation in frequencies not available to humans. They can "see" in the infrared and ultraviolet ranges or by sonar, radar, or lidar; they can "hear" microwave and AM andFM radio frequencies; they can "feel" using pressure sensors and temperature sensors; they can measure moisture content in the air or soil; and they can check acidity or alkalinity levels. The range of sensors available to robots is enormous—robots have access to much broader sensory information than those directly available to the human brain (humans can access the same hardware sensors, but the interfaces are somewhat different, and access to the information is likely much slower than for robots). Not only is the range of available sensors enormous, the range over which each sensor measures the particular phenomenon for which it was constructed is likely to be far greater than the human body is capable of. For example, the human body can sense only a (comparatively) narrow range of temperatures before the biological sensors in the body fail, or the brain instructs them to disengage, whereas robots are likely to be able to sense a far greater range before their temperature sensors start to melt or freeze.

An important question is how the different modes of embodiment and sensory capabilities affect how the human brain learns and perceives, versus how an ANN inside a robot learns and perceives. Observation would suggest the human brain/body combination has far more nuanced control over/reactions to sensory input—whether that's because the sensors have finer control and are more nuanced, or the brain's reactions, learned over time, are more nuanced, or, more likely, a combination of both, it is hard to gauge. A simple example is the difference between how a human approaches picking up a glass of water, especially for the first few times. Humans generally have far more finesse, much greater nuanced control, than do most robots, simply because most humans have learned over time how to approach such situations. ANNs can learn too, through feedback via robotic sensors, but humans have millions of years of learning pre-encoded in (the older parts of) their brains that they can draw on to assess both familiar and unfamiliar situations.

I think it is fair to say that although we can simulate, to some degree, the embodiment of a human brain and its immersion in its environment via its connection to the human body, how ANNs learn, how they develop "intelligence," from knowledge of their environment will be qualitatively different from the human brain.

Does Thought Require Language?

Language differentiates humans from all other animals. Language, at its simplest, is just a means of communication, but it is a special form of communication. Most, if not all, animals communicate, but only humans use language to convey information. Language implies syntax, in which word order determines meaning, and grammar, that defines rules to convey meaning. While some species communicate in sophisticated ways, some of which include vocalization, only humans use language with grammar and syntax (with the caveat that some species may have languages with syntax and grammar as yet undiscovered by humans). Moreover, every known human society has used language to communicate—language is ubiquitous among humans. All of this raises the question of how language might be related to the development and maintenance of human intelligence.

Poet and playwright, Oscar Wilde, wrote:[5] "It is only by language that we rise above them, or above each other—by language, which is the parent, and not the child, of thought" [20].

Polymath Bertrand Russell stated: "Language serves not only to express thoughts, but to make possible thoughts which could not exist without it" [21].

At face value both seem to suggest without language, thought is not possible.

Tufts University magazine, *Tufts Observer*, published an article by Tess Ross-Callahan that describes the experiences of a stroke victim left without language for a period of time, and Hellen Keller, who lost her sight and hearing at the age of 19 months, and did not begin to learn a language until she was almost seven years old [22].

In that article, Ross-Callahan reports the experience of neuroanatomist Dr. Jill Bolte Taylor who, after suffering a catastrophic stroke in 1996, lived for almost a decade without language—she didn't just lose the ability to speak, but actually had no access to the language she had learned throughout her lifetime. She had no words with which to form her thoughts (Bolte Taylor's own recounting of the story can be found in her 2008 book, *My Stroke of Insight* [23]). Bolte Taylor describes her experience during that almost 10-year period as being without words, without reflection. She says she experienced sensations (such as the sun shining), even emotions ("pure joy," in her words, in a 2010 interview with RadioLab), but is not clear on whether she experienced thought. Bolte Taylor says after some of her memories returned, she thought in images, rather than words.

Ross-Callahan quotes Helen Keller as saying: "When I learned the meaning of 'I' and 'me' and found that I was something, I began to think. Then consciousness first existed for me" [24].

---

[5] From Wilde's essay "The Critic as an Artist" (1891).

At face value this seems to suggest Keller believed that before learning a language, she was devoid of all thought. But in the *Tufts Observer* article, Ross-Callahan states: "It's not that Keller wasn't thinking before that day. She may have been volatile and violent, but she was able to identify people's faces with her touch, desire ice cream, and recognize repeated objects."

Indicating that whether or not Keller experienced direct thought, she did experience emotions and desires.

In her 1903 autobiography, Keller recalled the moment her teacher, Anne Sullivan, taught her the word "water," her first word: "I stood still, my whole attention fixed upon the motions of her fingers. Suddenly I felt a misty consciousness of something forgotten—a thrill of returning thought; and somehow the mystery of language was revealed to me" [25]. "Something forgotten," "a thrill of returning thought," is at least suggestive that Keller had experienced at least one thought to which she could return before learning language.

Only Hellen Keller and Dr. Jill Bolte Taylor really know what they experienced, but I don't see how it is possible to learn a language, to learn words and their meanings, without thinking. It may be that before learning (or re-learning/remembering in the case of Bolte Taylor) words and language, Keller and Bolte Taylor couldn't form complex thoughts, or carry on (mental) conversations with themselves, but the notion that they had no thought at all is, in my opinion, not reasonable.

Though language is important for developing thought, and especially engaging in communication and social interaction that may enhance thought, I do not believe language is a prerequisite for thought.

Can an ANN Match the Human Brain?
Might we one day construct supercomputers large enough to run ANNs with a similar number of neurons and connections as the human brain? Perhaps. Is it possible that in the future, we could construct those supercomputers small enough to rival the power requirements and physical size of the human brain, and mount them inside mobile enclosures with smart sensors? Perhaps. But the likelihood we could do either of those things any time in the foreseeable future, much less both, is vanishingly small.

For now we must be satisfied with the ANNs we can construct which, as we have seen, are orders of magnitude smaller in terms of the number of neurons and connections than the human brain. But the brain isn't a single, monolithic neural network. As previously described, the brain is composed of many interconnected neural networks, each (presumably) performing separate, specific functions—one neural network inside the human brain may perform its function and then pass its output(s) onto another neural network for further processing, in a divide-and-conquer fashion. The totality of those many biological neural networks, and their interaction, is, at least in part, what allows humans to display human intelligence.

Consider an area of the human brain known as the "fusiform face area" (FFA). The FFA, located in the inferior temporal cortex, is part of the human visual system and is thought to be one of the areas of the brain specialized for facial recognition [26]. The FFA is a localized group of interconnected neurons—a biological neural network. The human brain has several areas specialized for facial recognition (e.g. the occipital face area, OFA, and others) that work in concert to recognize faces in different circumstances, at different ages, etc. We already have very good facial recognition systems—machines utilizing ANNs specialized for facial recognition—and there is every reason to believe that we could combine several of those systems to construct a divide-and-conquer facial recognition machine that would rival the facial recognition ability of the human brain.

Alan Turing believed machines would, one day, think, and stated so in his 1950 paper: "…I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted."

The "end of the century" was, with hindsight, optimistic. I think it would be optimistic to suggest that we will be capable of constructing machines that could rival the human brain even in the next few decades. But I do think we could quite correctly claim that by using ANN technology we are even now able to construct machines that can match some specific functionalities of the human brain.

Human brains have been evolving for millions of years—hundreds of millions if we include our "reptilian brain" (the brainstem and the cerebellum). ANNs have been with us for only a handful of decades—and they are being helped along by a co-operative group of the most intelligent and complex thinking machines in the known universe. We should wait a few more decades before we dismiss ANNs as the basis of intelligent machines.

WILL MACHINES EVER DISPLAY HUMAN INTELLIGENCE?

Categorically yes. Some machines already do, or are very close—albeit in specific, constrained circumstances. Some examples are (not an exhaustive list):

- In 2018, Google Duplex, an AI-powered system that carries on human-like conversations, was widely reported as having passed the Turing test [27].
- Facial recognition software, coupled with state-of-the-art cameras, are very close to human standards.
- Some of the game-playing software now available, particularly games such as chess and Go, matches or exceeds human-level playing.
- AI-powered chatbots are closing in on human-level performance (such as IBM's Watson Assistant and Bold360 by LogMeIn).
- Intelligent automated systems (self-driving cars, home-automation systems, etc.) are very close to human standards, and will probably soon be ubiquitous.

- AI has been most prominent, and has produced very promising results, over the past decade or so in the field of data science ("Big Data"). The vast amounts of data now available, especially the extraordinary volumes of data collected by digital giants such as Google and Facebook, are almost impossible for humans to comprehend, let alone process in any reasonable timeframe.

In 2018 Gerard Doyle, network manager for Technology Ireland ICT Skillnet, asserted: "Many AI systems are displaying intelligence at a scale that is far beyond that of a human. It may be that in the future we will conclude that Turing picked the wrong benchmark—the human—and trying to prove that a machine can answer questions like a human may become an historical irrelevance" [28].

SOME THOUGHTS ON "THINKING LIKE HUMANS"

I stated earlier that what Hofstadter refers to as "mind, soul, 'I'" and the related concepts of consciousness, self-awareness, etc. were beyond the scope of this article, and that I would restrict this discussion primarily to issues surrounding machines displaying intelligent behavior. Now that the "primarily" part of the discussion is done, I will indulge in some speculation of some of the more esoteric philosophical issues—namely "mind, soul, 'I'" and the like.

The (original) question posed at the beginning of this article was "Will machines ever think like humans*?*".

I've already discussed, somewhat shallowly, what we mean by "thinking," but let's take that discussion a little further. What do we mean by "thinking like humans?" Humans are self-aware, conscious of their existence, of their "being." They have emotions—they are at different times happy, sad, angry, enraged, disappointed, elated, annoyed. They engage in free and original thought and daydreaming; they dream during their sleep; they have "aha!" moments. Humans are, or do, all those things, and more. Are those qualities and states a function of the brain and "thinking," or are they somehow separate? As I stated earlier, I think they are all emergent properties of the structure and function of the brain.

More particularly though, can a machine "think like a human" without displaying the qualities and states-of-mind discussed above? Can a machine think like a human, but not, for example, engage in free and original thought, even daydreaming, or be disappointed by something, or enraged about something? Can a machine that thinks like a human not be conscious and self-aware? Consider the reverse: Can we imagine a human that is not conscious of their own existence, that is not self-aware, that cannot engage in free and original thought, be said to be "thinking like a human?" Could they be said to be thinking at all? I don't think so, and I don't think the reverse is true. I think a machine that thinks like a human comes with all the qualities, emotions, and cognitive abilities (e.g. free and original thinking, daydreaming, etc.) that a thinking human does (and if it doesn't, it can't be said to think like a human).

How plausible is it that we could actually create machines that could engage in free and original thought? It's not so far-fetched. As way back as 1987, Erik T. Mueller published his Ph.D. thesis [29] "Daydreaming and Computation: A Computer Model of Everyday Creativity, Learning, and Emotions in the Human Stream of Thought," which was published as the book *Daydreaming in Humans and Machines* in 1990 [30]. The thesis and book describe the DAYDREAMER cognitive architecture developed by Mueller, which models the human stream of thought and its triggering and direction by emotions—effectively simulating human daydreaming.

Leaving aside for a moment the question of whether we can actually create machines that think like humans in the way just described, have we thought this through? Do we want to create machines that think like humans if they are going to come with all that that means? Let's think about that for a moment.

Why would we want to do that? We already have more than seven billion humans on the planet, many of whom do a perfectly satisfactory job of thinking like humans. Furthermore, we already know how to create more, and we have a good working pipeline already in progress, so waiting the 20 years or so it takes for the human brain to reach maturity shouldn't be a problem.

Are we hoping one day artificial intelligence will surpass human intelligence, perhaps even augment our own (via prostheses or implants), and lead us to places we couldn't go on our own?

Perhaps we think we need to do it—but do we? Do we really need machines that display human intelligence? Don't we just need algorithms—smart algorithms, maybe even intelligent algorithms—to allow machines to do the jobs for which they are intended (the jobs we can't, or don't want to, do)?

Or do we want to do it just to see if we can—to climb the metaphorical mountain just because it's there?

If we do create machines that think just like humans, that are self-aware and conscious of their existence, do we afford them the same rights as any other sentient being?

Are we ready—willing—to deal with issues attendant to creating machines that think like humans?

<p align="center">*****</p>

The following commentary is from members of the *Ubiquity* editorial board. Further commentary from interested readers is invited.

From **Ted Lewis**:

"You affirm your belief that AI can equal or surpass human intelligence without addressing the core issue of awareness and consciousness. Perhaps intelligence is not one thing. It might be levels. My dog shows signs of intelligence and self-awareness at one level while my 12-yr old shows signs at another level, etc. Most "intelligences" in your list are simply fast searches over big spaces that humans cannot compete with because of the relatively slow human brain. This seems to imply speed = intelligence, which is not cognition or consciousness."

Author's response:

I stated in the opening section of the article that I would not directly address consciousness and self-awareness. I'm not convinced either is a prerequisite for, or at the core of, intelligence—that would depend upon your definition of intelligence.

There may well be—probably is—an "intelligence gradient," or levels of intelligence. Solving most problems involves searching over large solution spaces looking for a good, if not "the best," solution—the ability to construct and search a large solution space quickly might well be evidence of intelligent behavior. All of these depend upon your definition of intelligence.

From **Peter Denning**:

"… consciousness and cognition … accompany intelligence and may even be preconditions."

"The things you list as recent developments foreshadowing machines that think are all simulations of intelligent behavior. Are simulations the same as the real thing?"

Author's response:

Do consciousness and cognition accompany intelligence? If they do in all examples we know of, must they? Are they preconditions? I suggest whether that proposition is true depends on your definition of intelligence.

I think that by considering consciousness and awareness to be at the core of intelligence, or that they accompany intelligence, you are assuming a definition of intelligence that implies that it (intelligence) is a quality or feature that is only present in sentient beings—or perhaps at a more basic level, beings that are conscious and self-aware. I think we can separate consciousness and awareness from intelligence —certainly from, in Turing's terms, the ability to display intelligent behavior (accepting that whether there is a qualitative difference between possessing intelligence and displaying intelligent behavior is liable to be contentious).

Whether simulations are the same as the real thing depends on your definition of intelligence and intelligent behavior. If by intelligent behavior you mean actual human intelligent behavior, then in a very narrow sense I agree, they are all simulations—how could they not be if they are not human? But if by intelligent behaviour you mean, in Turing's terms, behavior that is indistinguishable from human intelligent behavior (provided that behavior doesn't actually require flesh and blood…), then is it a simulation? The machine is displaying intelligent behavior

indistinguishable from human intelligent behavior—apart from humans being flesh and blood and machines being something else, how are the behaviors different? If I learn to do something that you can do, would you say I am simulating your behaviour, or copying it (and maybe even improving upon it)? Do we call it a simulation because humans came first? Isn't that just semantics? In the end, if the behaviour of the machine is indistinguishable from that of a human, does it matter what we call it?

From **Martin Walker**:
"Intelligence is exhibiting a faculty of understanding. When an intelligent agent has understood a situation or a phenomenon, then that agent is able to deal satisfactorily with the situation or phenomenon, independently of how, precisely, the situation or phenomenon is presented to the agent. This is manifestly not the case for artificial neural networks."

"The inferences made by ANNs are fragile and unstable (think, for example, of the image classifier that interprets a STOP sign as a speed limit sign after a few slight modifications to the original image)."

"I would suggest therefore that ANNs, as useful as they may be, do not exhibit intelligence."

Author's response:
Indeed, the "Chinese room" thought experiment was created to demonstrate John Searle's belief that exhibiting behaviour that one might initially believe to be intelligent does not imply understanding, and that intelligence without understanding is just an illusion. To buy into that argument we have to agree with Searle's belief that, in Martin's words, "exhibiting a faculty of understanding" is a necessary condition for intelligence. I am not convinced that it is—at least not until I know what is meant by "understanding." Neither, apparently, was Alan Turing—the Turing test makes no reference to understanding.

Consider an animal that is able to earn rewards, or elicit a predictable response from (say) humans, by learning that a particular sequence of actions on its part always results in rewards or a particular action by humans. The animal almost certainly doesn't understand why that happens—is it not exhibiting intelligent behaviour because it lacks that understanding? Or do we consider that simply understanding that there is a correlation between its behavior and a human's response (understanding at a level not too dissimilar from that of the translator in Searle's Chinese room) a sufficient prerequisite for intelligence? What is meant by "understanding"? What is the nature of "understanding" that would support a judgement of "intelligence?"

I know many people who learn things, even very complex things, by rote and never really understand them—how would I know (without testing them) by their behavior that they don't understand? Should I judge their behavior as any less intelligent than the person standing next to them that actually understands what they are doing?

22

These things depend on our definition of intelligence (and understanding). I think we're proving Sternberg's point: "Viewed narrowly, there seem to be almost as many definitions of intelligence as there were experts asked to define it."

While it is true that some Deep Neural Networks (DNNs) have been shown to be vulnerable to some input samples designed to fool the network [31], I think it is over-reaching to claim that "inferences made by ANNs are fragile and unstable"—that tends to imply that all ANNs are fragile and unstable by nature, which is not the case.

It is true that humans generally do a much better job of image recognition (probably through experience gained over a long period of time by seeing the object in question from many different angles and distances, obscured in various ways, etc., or by transferring generalizations learned from other, similar objects), but human brains are also vulnerable to some input samples designed to fool them—we call such input samples "optical illusions." Until a human brain has learned how to generalize and so recognize faulty/obscured (or otherwise adversarial) images (e.g. human infants), the inferences it makes are liable to be fragile and unstable.

ANNs are good at generalization—they don't have to be fragile and unstable. There are ways to mitigate adversarial input images (e.g. training a network with adversarial images), and work is being done in this area [32, 33, 34].

The human brain has many times the number of neurons and parallel pathways available to it than any ANN. My guess is that the human brain may have evolved multiple networks in the cortex to deal with input images, and that might allow it to be more robust when presented with adversarial images. As our ability to build faster ANNs with more nodes improves over time, we may find a similar approach helps ANNs deal with adversarial input images.

References

[1] Turing, A. M. Computing machinery and intelligence. *Mind* LIX, 236 (1950), 433–460. doi:10.1093/mind/LIX.236.433

[2] Searle, J. Minds, brains and programs. *Behavioral and Brain Sciences* 3, 3 (1980), 417-457. doi:10.1017/S0140525X00005756

[3] Commentary/Searle: Minds, Brains and Programs. *Behavioral and Brain Sciences*, 3, 417-457, 1980. doi:10.1017/S0140525X00005756

[4] Sternberg, R. J. *Beyond IQ: A triarchic theory of human intelligence.* Cambridge University Press, Cambridge, 1985.

[5] Gregory, R. L. *The Oxford Companion to the Mind*. Oxford University Press, Oxford, 1998

[6] Sternberg, R. J. Human intelligence. *Encyclopedia Britannica*. Dec 10. 2020; https://www.britannica.com/science/human-intelligence-psychology

[7] Wybo, W. A. M., Torben-Nielsen, B., Nevian, T., and Gewaltig, M-O. Electrical compartmentalization in neurons. *Cell Reports* 26, 7 (2019), 1681–1978. doi:10.1016/j.celrep.2019.01.074

[8] Hameroff, S. *Ultimate Computing*, Elsevier, 1987. ISBN 978-0444702838

[9] Penrose, R. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Oxford University Press, Oxford, 2016. ISBN 978-0-19-255007-1

[10] Hameroff, S., and Penrose, R.Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Neural Network World* 5 (1995), 793–804.

[11] Ghosh, S., Sahu, S., and Bandyopadhyay, A. Evidence of massive global synchronization and the consciousness. *Physics of Life Reviews* 11, 1 (2014), 83–84. doi:10.1016/j.plrev.2013.10.007

[12] Hameroff, S. and Penrose, R. Reply to criticism of the 'Orch OR qubit' – 'Orchestrated objective reduction' is scientifically justified. *Physics of Life Reviews* 11, 1 (2014), 104–112. doi:10.1016/j.plrev.2013.11.014

[13] Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., and Held, P. *Computational Intelligence: A Methodological Introduction*. Springer, 2013. ISBN: 978-1-4471-5012-1

[14] McCulloch, W. and Pitts, W. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics* 5, 4 (1943), 115–133. doi:10.1007/BF02478259

[15] Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. Cornell Aeronautical Laboratory. *Psychological Review* 65, 6 (1958), 386–408. doi:10.1037/h0042519

[16] Minsky, M. and Papert, S. *Perceptrons.* M.I.T. Press, Cambridge, 1969.

[17] Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79, 8 (1982), 2554–2558, doi:10.1073/pnas.79.8.2554

[18] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. Nature 323, 6088 (1986), 533–536, doi:10.1038/323533a0

[19] Larkum, M. E., Petro, L. S., Sachdev, R., Muckli, L. A perspective on cortical layering and layer-spanning neuronal elements. *Frontiers in Neuroanatomy* 12, 56 (2018). doi:10.3389/fnana.2018.00056

[20] Wilde, O. *Intentions*. James R. Osgood, McIlvaine and Co. 1891.

[21] Russell, B. *Human Knowledge: Its Scope and Limits*. George Allen & Unwin, Ltd. 1948.

[22] Ross-Callahan, T. The secret life of language: a world without words. *Tufts Observer*. Feb. 23, 2015; https://tuftsobserver.org/secret-life-of-language-2/

[23] Bolte Taylor, J. *My Stroke of Insight: A Brain Scientist's Personal Journey*. Viking, New York, 2008. ISBN 978-0-670-02074-4.

[24] Keller, H. *The World I Live In.* The Century Co.1908

[25] Keller, H. *The Story of My Life*. Doubleday, 1903.

[26] Kanwisher, N., McDermott. J., Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 11 (1997), 4302–11, doi:10.1523/JNEUROSCI.17-11-04302.1997

[27] Gewirtz, D. Google Duplex beat the Turing test: Are we doomed? ZDnet. May 14, 2018.

[28] Phillips, D. Would modern AI systems pass the 'Turing Test'? *The Irish Times*. Dec. 6 2018.

[29] Mueller, E. T. Daydreaming and Computation (Technical Report CSD-870017). 1987.

[30] Mueller, E. T. *Daydreaming in Humans and Machines*. Ablex Publishing Corp. 1990. ISBN:978-0-89391-562-9

[31] Heaven, D. Why deep-learning AIs are so easy to fool. Nature 574 (2019), 163–166, doi: 10.1038/d41586-019-03013-5

[32] Ren, K., Zheng, T., Qin, Z., and Liu, X. Adversarial Attacks and defenses in deep learning. *Engineering* 6, 3 (2020), 346–360. doi: 10.1016/j.eng.2019.12.012

[33] Yuan, X., He, P., Zhu, Q., and Li, X. Adversarial examples: attacks and defenses for deep learning. *IEEE Transactions on Neural Networks and Learning Systems* 30, 9 (2019), 2805_2824, doi: 10.1109/TNNLS.2018.2886017

[34] Akhtar, Z. and Dasgupta, D. A brief survey of Adversarial Machine Learning and Defense Strategies. The University of Memphis. 2019. Technical Report No. CS-19-002

Biography

Jeff Riley is a 40-year veteran of the computing industry, and has worked in many roles in the field including Master Technologist and Scientist with Hewlett-Packard. Jeff holds a master's degree in IT, a Ph.D. in AI, and is a former Adjunct Principal Research Fellow of RMIT University in Melbourne, Australia. He is the founder of Praescientem, a company specializing in AI and IT consulting and education services. Jeff is currently pursuing a second Ph.D., in theoretical astrophysics, and builds robots in his spare time. Jeff's major areas of interest in the AI space are machine learning and evolutionary computation. His publications are available at https://www.praescientem.com.au/pages/research.html.