After All, Artificial Intelligence is not Intelligent: in a Search for a Comprehensible Neuroscientific Definition of Intelligence

Received: January 7th, 2022 • Approved: November 16th, 2022
https://doi.org/10.22395/ojum.v21n46a9

Sthéfano Divino
Centro Universitário de Lavras (Unilavras), Lavras, Brazil
sthefanoadv@hotmail.com
https://orcid.org/0000-0002-9037-0405

ABSTRACT

This paper explores a series of thoughts about the meaning of intelligence in neuroscience and computer science. This work aims to present an understandable definition that fits our contemporary artificial intelligence background. The research methodology of this essay lies in existing theories of artificial intelligence, focused on computer science and neuroscience. I analyze the relationship between intelligence and neuroscience and Hawkin's Thousand Brains Theory, an approach to show what it is an intelligent agent according to neuroscience. Here, the main result relies on the verification that intelligence is only possible in the neocortex. According to this result, the study performs a second critical analysis aiming to demonstrate why there is no artificial intelligence today.

Keywords: artificial intelligence; computer science; intelligence; machine learning; neuroscience.

Al fin y al cabo, la inteligencia artificial no es inteligente: en la búsqueda de una definición neurocientífica comprensible de la inteligencia

RESUMEN

Este trabajo explora una serie de reflexiones sobre el significado de la inteligencia en la neurociencia y la informática. El objetivo de este trabajo es presentar una definición comprensible que se ajuste a nuestro entorno contemporáneo de inteligencia artificial. Se analiza la relación entre la inteligencia y la neurociencia y presento la teoría de los mil cerebros de Hawkins, un enfoque para mostrar qué es un agente inteligente según la neurociencia. Aquí, el principal resultado se basa en la comprobación de que la inteligencia sólo es posible en el neocórtex. De acuerdo con este resultado, el estudio hace un segundo análisis crítico con el objetivo de demostrar por qué no existe la inteligencia artificial en la actualidad. La metodología de investigación de este ensayo se basa en las teorías existentes sobre la inteligencia artificial, centradas en la informática y la neurociencia.

P*alabras clave*: inteligencia artificial; informática; inteligencia; aprendizaje automático; neurociencia.

Afinal de contas, a inteligência artificial não é inteligente: à procura de uma definição neurocientífica compreensível da inteligência

RESUMO

Este trabalho explora uma série de reflexões sobre o significado da inteligência na neurociência e informática. O objetivo desse trabalho é apresentar uma definição compreensível que se ajuste ao nosso ambiente contemporâneo de inteligência artificial. Analisa-se a relação entre inteligência e a neurociência e a teoria dos mil cérebros de Hawkins, uma abordagem para mostrar que é um agente inteligente segundo a neurociência. O principal resultado se baseia na comprovação de que a inteligência só é possível na neocórtex. De acordo com esse resultado, o estudo faz uma segunda análise crítica com o objetivo de demonstrar por que não existe inteligência artificial na atualidade. A metodologia aplicada a esta pesquisa baseou-se nas teorias existentes sobre a inteligência artificial, centradas na informática e na neurociência.

P*alavras-chave*: inteligência artificial; informática; inteligência; aprendizagem automática; neurociência.

Introduction

This paper is a part of my main study developed in my Ph.D. dissertation, where I want to show how artificial intelligence agents can or cannot have legal personhood, his (un) consequences, and how this should be possible by analyzing Legal Theories. According to Russell & Norvig (2010, p. VIII) AI means "agents that receive percepts from the environment and perform actions". In the same meaning, Franklin & Graesser (1997, pp. 21-35) postulates that "An autonomous agent is a system situated within and a part of an environment that senses that environment and acts on it, over time, in pursuit of its own agenda and so as to effect what it senses in the future". Furthermore, "autonomous agents possess goals which are generated from within rather than adopted from other agents. These goals are generated from motivations which are higher-level non-derivative components characterizing the nature of the agent, but which are related to goals" (Luck & d'Inverno, 1995, p. 254-260). The question that guides this work is: what is the meaning of intelligence? This question does not arise just if an artificial agent is intelligent, but in a human too. This essay is a fraction of this whole work. Here, I bring the results of one of my chapters.

It may be easy to get an idea of how the human brain works, but neuroscientists are in the dark when they open this pandora's box. To understand this scenario, I choose to use Hawkins and his thousand-brain theory. The results obtained by the author in years of research demonstrate through his theory how the brain can function and, consequently, what intelligence is. His conclusions appear logical and verifiable, and with this, I can bring considerable reflections on how machines can or cannot be responsible and have rights and duties.

My claim is to show what is intelligence (in its neuroscientific meaning). It is the modular ability to learn the world, and that learning occurs in the neocortex through the association between neurons, cortical columns, and reference frames. Thus, an artificial agent must at least have the same qualities as the brain to be intelligent. This proposition seems intelligible because the brain is the only reference to what we understand as intelligent. According to Hawkins, I can suggest that artificial agents cannot be intelligent (yet) due to the gap between neuroscience knowledge and Machine Learning development.

This paper has two sections. In the first section, I analyze the relationship between intelligence, neuroscience, and present Hawkin's Thousand Brains Theory, an approach to show what it is an intelligent agent according to neuroscience. In the second section, supported by Hawkins's theory, I want to show why there is no artificial intelligence today. For this, the research methodology of this essay lies in existing theories of artificial intelligence, focused on computer science and neuroscience.

2. Intelligence, Neuroscience and Hawkin's Thousand Brains Theory

Computer science sees artificial intelligence as an objective criterion (Shrestha, Ben-Menahem, & Von Krogh, 2019)[1]. Between these objective criteria is the reason. According to Haggard (2017, p. 196-207), "reason is the Sense of agency that refers to the feeling of controlling one's own actions and, through them, events in the external world". Thus, intelligence in computational terms is closely related to reason. However, a rational agent does not always act in a 100 % rational way. If you ask a child how much is 2+2 he or she would proudly answer "5", his or her answer is autonomous but irrational. Even if one tries to demonstrate the result of 4 by putting 2 bananas in the basket and then two more bananas in the same basket, the child may ignore this fact and persist that the answer is, in principle, according to his or her understanding, five. The connection between the terms agent and environment is evident, but also rationality cannot be ignored. Therefore, one cannot be defined without the presence of the other. According to Sutton & Barto (1998),

> The reinforcement learning problem is meant to be a straightforward framing of the problem of learning from interaction to achieve a goal. The learner and decision-maker is called the agent. The thing it interacts with, comprising everything outside the agent, is called the environment. These interact continually, the agent selecting actions and the environment responding to those actions and presenting new situations to the agent (p. 53).

Another example of this situation is when human beings make decisions based on emotions. Therefore, the concept of intelligence despite being tied to reason is not limited to it. Some intelligent agents do not act rationally. Computer science requires an interaction between the agent and the environment in which it is located to be an intelligent agent (Bostrom & Yudkowsky, 2011; Bostrom, 2014). There must be political and social factors that affect decision-making patterns.[2] AI must be connected to the edges of these structures to escape a strictly technical definition (Brayne, 2021; Crawford, 2021). Thus, the objective concept of intelligence is not just a singular dimension or concept, but a rich space structured in information processing capabilities. In this contextualization, an AI can supposedly act autonomously because of its intelligence. The autonomy of an AI lies in the degree of sophistication and human intervention in its conduct. Thus, agency or self-determination in its compu-

[1] "The fundamental idea of decision theory is that an agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes MAXIMUM EXPECTED of the action. This is called the principle of maximum expected utility (MEU). Note that UTILITY "expected" might seem like a vague, hypothetical term, but as it is used here it has a precise meaning: it means the "average," or "statistical mean" of the outcomes, weighted by the probability of the outcome" (Russell & Norvig, 2010, p. 483).

[2] "A motivation is any desire or preference that can lead to the generation and adoption of goals and that affects the outcome of the reasoning or behavioral task intended to satisfy those goals. An autonomous agent is an agent with a non-empty set of motivations". (Luck & d'Inverno, 2001, pp. 1-20). For more, see Beer, Fisk & Rogers (2014), and Huang et al. (2007).

tational conception, therefore, on an abstract level would be possible, distinguishing itself through criteria and decision-making skills of machines from those of human beings. These considerations allow us to consider the potential for an AI to act autonomously and realize its agency in the factual environment where it is executed, especially if the correct ML models are adopted for this task.

These propositions, so far, are limited to a panorama of computer science that bases its notion of intelligence on the mammalian brain. This is so because the brain is the only thing that we know so far to be intelligent. But for neural networks to be equal to the human brain, we must understand how it works. In the realm of computer science, this understanding is wrong. So, to verify the propositions above I will try to describe how the brain works and if AI is intelligent according to neuroscience.

Since this is a complicated theme to define, I am looking for more adequate and updated studies that address in a verifiable way how intelligence can or could/should be understood according to contemporary technical and scientific advances. Therefore, I do not make philosophical reflections (respecting them in their limits, at this moment), but I look for studies with practical and concrete results, even theoretically synthesized, about the functioning of the brain and intelligence.

It may be easy to get an idea of how the human brain works, but neuroscientists are in the dark when opening pandora's box. I choose to use Hawkins's Thousand-Brain Theory to understand this scenario. The results obtained by the author in years of research demonstrate through his theory how the brain can function and, consequently, what intelligence is. His conclusions appear to be logical and verifiable, and with this I can present paths for a feasible understanding of intelligence.

2.1 The early paths of Intelligence in Neuroscience

Understanding intelligence goes through a path of overcoming. Since the 1970s, when Francis Crick (1979, pp. 219-233) wrote an article called Thinking About the Brain, considerable data about brain works were already available. For Crick, the brain was like a puzzle where we have a reasonable amount of data, but we don't know how these data are connected and interconnected among themselves (Hawkins, 2021, p. 16). One of the problems arising from this correlation is how intelligence is made in the brain.

Hawkins (2021) notes that to understand how intelligence is made is necessary to locate the appropriate brain section for such a task. That is: where is intelligence produced? According to Hawkins (2021, p. 25)[3], the neocortex, present only in mammals, is the region responsible for the reserve and production of all human intelligence. "Almost all the capabilities we think of as intelligence —such as vision, language, music, math, science, and engineering— are created by the neocortex. When we think about

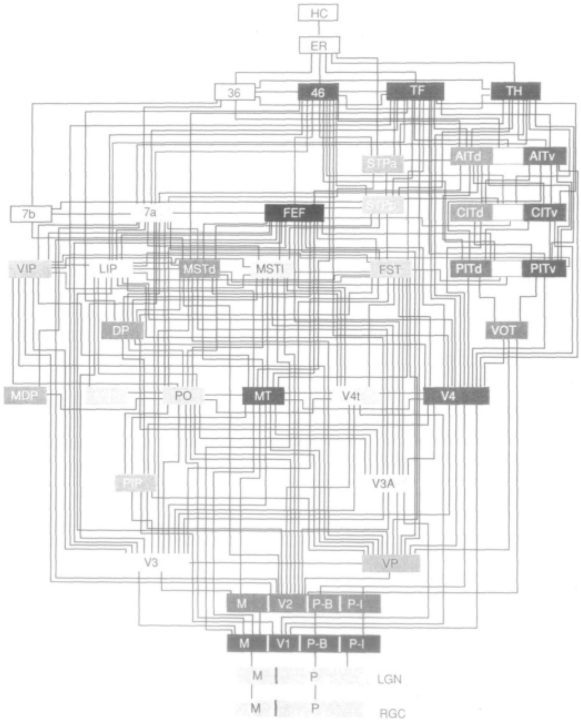
[3] Hawkins is not alone. Other neuroscientists defend this idea.. See: (Lindenfors, 2005), (Barton, 1996), and recently (Michaud, 2016).

something, it is mostly the neocortex doing the thinking" (Hawkins, 2021, p. 25). Hawkins understands Intelligence as the capacity to understand the world, such as the shape of objects; the behavior of creatures; how doors open and close; and our location and position in the world as human beings.

Interestingly the neocortex alone does not exert direct control over behavior or muscle movements (Yokoi & Diedrichsen, 2019). Breathing and other basic movements will be made by the old brain, which is also responsible for the most primitive functions of human beings, such as survival, sex, and proliferation. But unlike the brainstem, where there is an almost perfect cross-section of its parts, the neocortex has different zones of action, with their respective subdivisions, but there is no possibility of visualizing them clearly as in the brainstem.

For example, Felleman and Van Essen (1991) show how the neocortex of a primate works, where nerve synapses were detected from stimuli made by the researchers. Each rectangle in the image represents different regions of the neocortex, and the lines represent the direction of the nerve synapses and how information travels from one side to another.

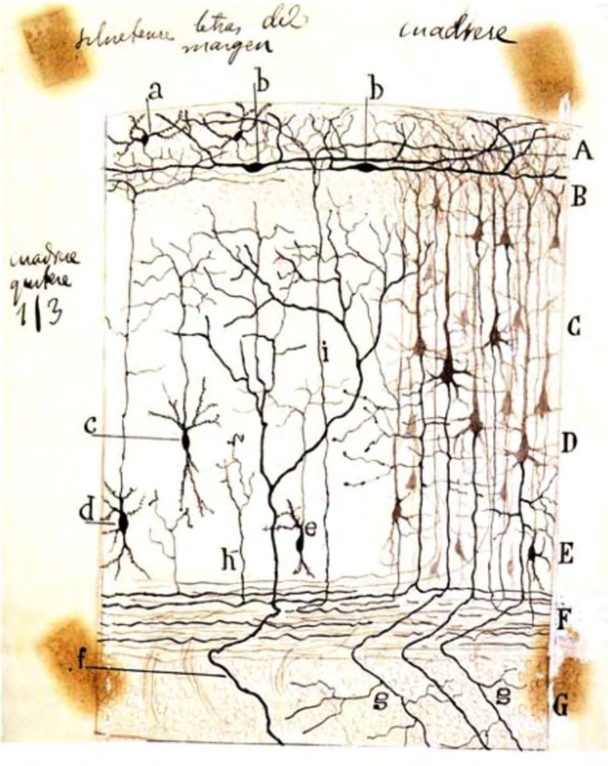
Figure 1. Connections in a Primate's Neocortex



S*ource*. Felleman and Van Essen,1991.

We can have an idea that the working of the neocortex is similar to that of a flowchart whose interpretation would be based on the hierarchy and positioning of each brain zone. However, the image itself shows that the flow can go either horizontally or vertically. Although its external appearance is uniform, Ramón y Cajal (1923), a neuroscientist of the 1800s, was responsible for demonstrating how the nerve cells present themselves in the neocortex (Gil *et al.*, 2014). The result of his research, which can be seen below, earned him the Nobel Prize.

Figure 2. A section inside of a little mammal neocortex



S*ource*. Ramón y Cajal (1923).

The results brought by Ramón y Cajal showed that the neurons present in the neocortex are arranged in six layers (Ogawa *et al.*, 1995), as well as having axons and dendrites, responsible for the nerve connections called synapses (Hawkins, 2021, p. 29). But this observation can only be made if seen from the inside. The neocortex looks the same and its regions cannot be distinguished from the outside. But what is the reason? Edelman and Mountcastle (1978) proposed that its similarity is because the whole neocortex is performing the same activity simultaneously, so there would be no reason to differentiate it. In this way, by discovering how a small region works, we could discover how the whole system works as well (Hawkins, 2021). But the authors' main merit was to verify that intelligence can manifest itself as an algorithm in the cerebral cortex through a cortical column.

> Cortical columns are not visible under a microscope. With a few exceptions, there are no visible boundaries between them. Scientists know they exist because all the cells in one column will respond to the same part of the retina, or the same patch of skin, but then cells in the next column will all respond to a different part of the retina or a different patch of skin. (Hawkins, 2021, p. 37)

Edelman & Mountcastle (1978) propose that the neocortex has about 150,000 cortical columns. The assumption is made because they cannot be seen even with microscopes. Hawkins says that scientists know that they exist because of the response of other cells to stimuli made between them, but they remain a mystery in the realm of neuroscience.

2.2 A Model of World in Your Head: Cortical Columns and Reference Frames

Hawkins' assumptions, however, become necessary because it is precisely in this place where the development of intelligence is assumed. Our brain reacts to stimuli made by sensors (hands, eyes, mouth, etc.) to produce expected and unexpected results. These results, however, are not random but have a correction between what you and I expect to see and what you and I expect to feel when we visualize and touch something. Every time neocortex makes these predictions to understand and map the world around it. It is from this context that the brain learns what is normal, what is abnormal, and what can be expected based on experience. As a result, you get what Hawkins (2021) calls A Model of World in Your Head.

> The brain creates a predictive model. This just means that the brain continuously predicts what its inputs will be. Prediction isn't something that the brain does every now and then; it is an intrinsic property that never stops, and it serves an essential role in learning. (Hawkins, 2021, p. 42)

If a prediction is made erroneously, the brain will try to make the correction and conform to the expected or, as a last resort, learn not to replicate the behavior as a consequence (Konflanz, 2019); (Byrne & Corp, 2004); (Deaner, Isler, Burkart & Van Schaik, 2007).

A newborn baby does not have any kind of experience or model of the world in its neocortex. "When you are born, your neocortex knows almost nothing. It doesn't know any words, what buildings are like, how to use a computer, or what a door is and how it moves on hinges. It has to learn countless things" (Hawkins, 2021, p. 44). It is through its development and the acquisition of experience that a model of the world will develop in the neocortex of each individual. This model will be responsible for actions, predictions, and perceptions. From this observation, Hawkins draws two observations: 1) the world is constantly changing and, consequently, intelligent beings must adapt to this change; and 2) as we move, our conception and model of the world change every time we move around because brain inputs provide us with new data and information to implement the experience so far lived. It is in this way that sensory-motor learning develops.

According to Hawkins (2021), the brain learns a model of the world through our observation, a result of how our senses and sensors interact with the world and the information that is perceived by us. From this assumption arise two principles of neuroscience: Principle number one: thoughts, ideas, and perceptions are the activity of neurons, and; Principle number two: everything we know is stored in the connections between neurons. When we learn something new, connections between neurons are reinforced. When we forget something, these connections are weakened, and, interestingly, new synapses are not only (re)forced, but also created and undone with time and experiences acquired by the individual (Li *et al.*, 2017; Kim and Thayer, 2001).

For these reasons, understanding the brain as a machine or a computer does not seem intelligible. The neocortex responsible for learning the world as a structured model becomes an adequate idea due to its verification. Hawkins demonstrates that the brain can make predictions and forecasts due to the existence of two types of neurons: neurons that fire when the brain is seeing something, and neurons that fire when the brain is predicting that it will see something. The brain needs to keep its predictions separate from reality to avoid hallucinations. Using two sets of neurons does this very well. However, for Hawkins, there are two problems with this idea.

> First, given that the neocortex is making a massive number of predictions at every moment, we would expect to find a large number of prediction neurons. So far, that hasn't been observed. Scientists have found some neurons that become active in advance of an input, but these neurons are not as common as we would expect. The second problem is based on an observation that had long bothered me. If the neocortex is making hundreds or thousands of predictions at any moment in time, why are we not aware of most of these predictions? If I grab a cup with my hand, I am not aware that my brain is predicting what each finger should feel, unless I feel something unusual—say, a crack. We are not consciously aware of most of the predictions made by the brain unless an error occurs. Trying to understand how the neurons in the neocortex make predictions led to the second discovery. (Hawkins, 2021, p. 52)

Furthermore, Hawkins (2021) finds that predictions are made within neurons. Since predictions can appear in two ways, one while the world is changing around us and the other because we are moving with the world, neurons need to figure out how much context is needed to make proper predictions.

> A prediction occurs when a neuron recognizes a pattern, creates a dendrite spike, and is primed to spike earlier than other neurons. With thousands of distal synapses, each neuron can recognize hundreds of patterns that predict

> when the neuron should become active. Prediction is built into the fabric of the neocortex, the neuron. (Hawkins, 2021, p. 57)

But prediction is an ambiguous function of the brain, and neuroscientists know this. Yet even in simulated tests, Hawkins was able to prove that neurons can learn diverse sequences and memorize them. Even with the death of about 30 % of them, the sequence was still memorized by the rest (Hawkins & Ahmad, 2016). The results demonstrate that predictions can happen inside neurons.

But how does access to this information happen? As visualized before, it seems to exist, in an intelligible and acceptable way, a map (analogically) in our heads. For each area of the neocortex, we can fraction it until we get to the cortical columns, which must be understood as referential frames that can respond to input through experience and the stimuli that are presented to it. Our brain doesn't process a picture or an image: its comprehension starts from the reception by sensors behind the eyes, which are then divided into considerable parts. After reaching its destination, we can understand the location, shape, and even texture of the object observed. This is one of the primary functions of the neocortex according to Hawkins: the processing of reference frames.

But why are Hawkins' reference frames important for understanding intelligence? First, they allow the brain to grasp the structure of something. Second, by defining an object using a reference frame, the brain can manipulate it entirely at once. Third, referential frames are necessary for planning and creating movements. Referential frames are used in many fields. Scientists working in the robotics program, imagine and plan what the movements of a robot's arm or body should look like. Reference frames are also used in animated movies to render characters as they move.

With this result, Hawkins was able to assume the difficulty of understanding how the apprehension of meaning in the neocortex takes place. However, as hard as the concept may be, its meaning must be gradual and stored in different points of the brain. Thus, when someone wants to understand the concept of democracy, we must review the concept of State and Government, for example. Each of these concepts is located in a reference frame and, when connected, can build other reference frames aimed at understanding the concept of democracy. The same happens with objective good faith, a concept that presupposes the existence of other reference frames, such as principles, norms, laws, contracts, property, and processes. Each of these assumptions is stored in different referential frames that, when activated by the neocortex, its synapses snap together and build individualized concepts from its previous knowledge. In other words, Hawkins demonstrates that a good background is intrinsically linked to data and facts (Hawkins, 2021, p. 86).

The problem is accentuated when language is included as one of the most important cognitive abilities capable of distinguishing human beings from other animals. Hawkins claims that without the ability to store and share knowledge and experience through language, modern society would be impossible to realize. According to Hawkins, Etard *et al*. (2000) and Naeser *et al*. (1987), there are two areas in the neocortex of considerable size that are responsible for language[4]. The section responsible for language comprehension is Wernicke's, while the section responsible for its production is Broca's. However, there is no consensus on the exact location and extent of each of these regions. Hawkins also points out that these areas are not necessarily differentiated between comprehension and production, since both can act interdependently and simultaneously with similar results, and that language cannot simply be isolated in two small regions of the neocortex.

It is the experience, behavior, and contact with the world that will be able to fill the frames of reference and create new neurological connections to enable the intelligibility of language and, consequently, to disseminate knowledge. Every moment that information enters through observation and body sensors, the world in our brain undergoes a drastic change: new frames of reference are made by new synapses, and new synapses are undone. Although each cortical column doesn't necessarily know what they are learning and also doesn't know what these models represent, individually they don't make sense, when put into a referential frame they begin to make sense and demonstrate how the brain works through them. It is at this point that Hawkins' Thousand Brain Theory is developed and presents a seemingly acceptable answer to the AI problem.

2.3. Hawkins' Thousand Brain Theory

An attempt is made to overcome the idealization of the brain as a flowchart and present it as referential frames. The cortical columns, even without less sensitive levels, can grasp and recognize objects. Thus, the neocortex has several models of a particular object. This model can be in different referential frames, which are not identical, but complementary. This complementarity is essential because according to Hawkins a cortical column cannot learn a referential model of every object in the world. Such a situation would be impossible because there is (yet) an unknown limit to how many objects each one could learn. But when they are connected, even if in different regions, they can recognize what is intended by the brain by accessing the previously learned model.

Thus, knowledge for Hawkins is distributed in the brain. Nothing we know is stored in just a single place, a single cell, or a single frame referent. Knowledge is stored everywhere and distributed in multiple columns. Everything in the brain works interdependently. A neuron does not depend on just a single synapse, but on about thirty to recognize a pattern.

[4] However, on the other side, see: (Binder, 2017).

> Therefore, we should not be surprised that the brain does not rely on one model of anything. Our knowledge of something is distributed among thousands of cortical columns. The columns are not redundant, and they are not exact copies of each other. Most importantly, each column is a complete sensory-motor system, just as each water department worker is able to independently fix some portion of the water infrastructure. (Hawkins, 2021, p. 98)

The problem lies in ascertaining how this information is linked and tied together. Hawkins (2021) proposes the question: how are our sensory inputs tied to a singular perception? There is no randomness, but a voting system elaborated by Hawkins that accesses information according to its sense and intelligibility and brings, at least, un-derstanding to the individual. If a person inserts his or her hand into a black box with an object inside and begins to feel it, the nerve synapses will begin to call each other and "vote" to determine which object among those you have already touched it best represents. It can be a cup, a glass, a pen, or any other physical object. If you have had previous contact, the neocortex will start and finish this vote-counting system to determine which object is the one touched and felt.

Since the connections in each column go up and down using the layers, remaining largely within the boundaries of each column, Hawkins states that only a few cells will be able to vote. Most of them do not represent any kind of information that the others could opine on. It is not a complicated view, but I prefer to keep his words and his original meaning:

> Using its long-range connections, a column broadcasts what it thinks it is observing. Often a column will be uncertain, in which case its neurons will send multiple possibilities at the same time. Simultaneously, the column receives projections from other columns representing their guesses. The most common guesses suppress the least common ones until the entire network settles on one answer. Surprisingly, a column doesn't need to send its vote to every other column. The voting mechanism works well even if the longrange axons connect to a small, ran-domly chosen subset of other columns. Voting also requires a learning phase. In our published papers, we described software simulations that show how learning occurs and how voting happens quickly and reliably. (Hawkins, 2021, p. 104)

With this proposal, Hawkins (2021) solves another mystery of the human brain: how our perception of the world seems to be stable when our brain is constantly changing in the face of incoming information (Tehovnik & Slocum, 2004). As low as the number of active neurons is extremely low (about 2 % only), they are responsible for understanding the world and maintaining stability in our brain. But no matter how small your attention is directed to an object, event, or person, the neocortex never stops creating learning models, and it is these learning patterns that are ephemeral and long-lived.

Hawkins' theory seems to be correct and, as much as it is not, it is testable and verifiable (and this is important). Notice that when we deal with the brain and intelligence in its neurobiological terms there is a substantial differentiation from the Neural

Network of the previous chapter. Even in a skeptical view, it seems that the computational approach is even utopian and fictitious.

The intelligence assumptions developed by Hawkins are tied to frames that are grasped in a much more complex way than artificial neural networks. Thus, if his theory is correct, I suggest that the future of artificial intelligence will be substantially different from what we have seen before, and what most computer scientists predict. The stance is skeptical, but with a view based on the science developed so far. And it is for this reason that we must analyze whether an AI is intelligent.

3. Why is there no Artificial Intelligence (yet)? The Problem of knowledge Representation

We can see that the biggest reason why artificial intelligence is not intelligent is that it can only do one or a few things, while humans can do several. In this way, artificial intelligence systems are inflexible. Any human being can learn through new experiences: even if it is a hard task and we are good at some and not at others, the possibility exists. Farming, programming, piloting, driving, composing, creating, and manufacturing are just some of the hundreds of thousands of skills that we can develop during our lifetime. Systems that operate in ML do not have this flexibility. An AI can beat the best chess player in the world, but it doesn't know how to do anything else but play chess.

Computer science seeks to develop machines that are equal to human intelligence and consequently learn new tasks, assimilate them with other activities, and are flexible enough to solve new problems. This is where AGI comes in. For Hawkins, however, the path taken so far does not lead us to AGI, because the ML system does not put us in a suitable place to create truly intelligent machines. For the author, we cannot achieve AGI by doing little more than we already are. We would need a different approach.

It is not enough to follow the path that machines should outperform humans in specific tasks (although they can). I must focus on the learning flexibility of AI so that it performs better than humans and achieves a result of being able to perform and learn as many things as possible just like humans. This learning and development of everyday knowledge are easy for us humans. But when viewed from a programming perspective, the scientists responsible for developing AI have not yet figured out how to do this with a computer. No matter how much neural networks exist, and the construction of software structured in schemas and frameworks intended for the organization of the knowledge learned so far, the result is still very uncertain, unintentional, and often unexpected. The world has a complexity that cannot be summarized in an algorithm. The amount of things that a child can know and learn seems to be impossible to verify and reduce in AI programming. This is the problem of knowledge representation in an AI (Markman, 2013).

> What is a knowledge representation? We argue that the notion can best be understood in terms of five distinct roles that it plays, each crucial to the task at hand:

> First, a knowledge representation is most fundamentally a surrogate, a substitute for the thing itself, that is used to enable an entity to determine consequences by thinking rather than acting, that is, by reasoning about the world rather than taking action in it. Second, it is a set of ontological commitments, that is, an answer to the question, In what terms should I think about the world? Third, it is a fragmentary theory of intelligent reasoning expressed in terms of three components: (1) the representation's fundamental conception of intelligent reasoning, (2) the set of inferences that the representation sanctions, and (3) the set of inferences that it recommends. Fourth, it is a medium for pragmatically efficient computation, that is, the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance that a representation provides for organizing information to facilitate making the recommended inferences. Fifth, it is a medium of human expression, that is, a language in which we say things about the world" (Davis *et al.*, 1993, p. 17).

Hawkins points out that knowledge representation is not only one of the problems of AI but the only problem. As advanced as neural networks are, they are still insufficient to develop knowledge. Even though scientists try to train them on a base and a database, programming by DL is not sufficient to shape knowledge just as a child develops when learning new skills. There is only a representation and copying of considerable statistics coming from the countless data into its software.

In this sense, Hawkins sees that the only way to develop an AI would be initially understanding how the brain works since it is the only thing, we know that is intelligent. However, for the author, we are still far from programming AIs close to the brain's functioning, because it is not enough just to understand how the brain works, although this is only the first step towards creating an AI.

Hawkins uses an analogy to demonstrate how the Thousand Brain Theory solves the problem of knowledge representation:

> The Thousand Brains Theory solves the problem of knowledge representation. Here is an analogy to help you understand how. Let's say I want to represent knowledge about a common object, a stapler. Early AI researchers would try to do this by listing the names of the different parts of the stapler and then describing what each part does. They might write a rule about staplers that says, "When the top of the stapler is pressed down, a staple comes out of one end." Words such as "top," "end," and "staple" had to be defined to understand this statement, as did the meaning of the different actions such as "pressed down" and "comes out." And this rule is insufficient on its own. It doesn't say which way the staple faces when it comes out, what happens next, or what you should do if the staple gets stuck. So, the researchers would write additional rules. This method of represen-ting knowledge led to a never-ending list of definitions and rules. AI researchers didn't see how to make it work. Critics argued that even if all the rules could be specified, the computer still wouldn't "know" what a stapler is.
>
> The brain takes a completely different approach to storing knowledge about a stapler: it learns a model. The model is the embodiment of knowledge. Imagine

> for a moment that there is a tiny stapler in your head. It is exactly like a real stapler —it has the same shape and the same parts to move in the same way— it's just smaller. The tiny model represents everything you know about staplers without needing to put a label on any of the parts. If you want to recall what happens when the top of a stapler is pressed down, you press down on the miniature model and see what happens. (Hawkins, 2021, pp. 122-123)

In an objective sense, knowledge is a model structured in cortical columns. For an AI to be intelligent and to reach the level of AGI, its constructions must perform just as the human brain does: using referential models, as if they were maps, to understand the world just as it happens in the neocortex. Without this assumption, Hawkins does not see a possibility of creating a real intelligent AI.

I must agree with Hawkins. No matter how universal neural networks are presented (as proposed by Gharbi, Elsharkawy & Karkoub, 1999; Qiu *et al.*, 2020; Sun & Ertekin, 2015), we must recognize their limitations and verify that they cannot multitask like a human being. While a child can interact, run and talk, as well as an adult can possess the ability to drive a car and fire a perfectly good shotgun, an artificial intelligence founded on a neural network does not even come close to this complexity. An AI is necessarily trained to outperform humans in a single task, but the point of AGI is to have a machine equivalent to humans in all or most tasks. Thus, for an agent to be considered intelligent it must have a brain-based model, as this is the only thing we have known as intelligent so far.

Hawkins presents four attributes to verify that an AGI could be considered intelligent. The first is continuous learning. Since we are learning during life, there is no way to expect less from an AI. The world changes constantly and there is a need to assimilate this change with our life claims to reflect in the claimed results. Most AI systems today do not exhibit this quality. They are trained on a database and when their training is complete, they are untrained and removed from that place. Thus, there is no such flexibility criterion.

Even if we are talking about supervised or unsupervised AI activities, as well as AI developed in multilingual ML systems, the results are inflexible compared to that of the brain. When a neuron learns a new world pattern, new shapes and synapses appear inside it. These synapses do not affect the previously learned. Thus, learning something new does not force the subject to modify or forget something previously learned. Artificial neural networks today do not yet possess this ability.

Moreover, since intelligence requires learning by modeling the world, we cannot view it as if it were something singular. As we move and move, intelligence and knowledge are created and developed. For Hawkins, movement is indispensable in the task of learning. Human beings cannot learn new models of objects without contacting and interacting with them. These movements don't need to be physical; it is enough for the brain to understand a change in the environment and in the inserted information (such as concepts) for effective learning to occur. From these referential models established through behavior and nerve synapses, prediction can happen, and intelligence will be taken as the expected result.

From the set of several models, which represents the third attribute of Hawkins, the neocortex will distribute the intelligence in several of them to create certain flexibility in the access to information. And it will be through the reference frames (fourth attribute) that the knowledge generated by the previous processes will be stored.

> In the brain, knowledge is stored in reference frames. Reference frames are also used to make predictions, create plans, and perform movements. Thinking occurs as the brain activates one location at a time in a reference frame and the associated piece of knowledge is retrieved.
>
> Why is it important? To be intelligent, a machine needs to learn a model of the world. That model must include the shape of objects, how they change as we interact with them, and where they are relative to each other. Reference frames are needed to represent this kind of information; they are the backbone of knowledge.
>
> How does the brain do it? Each cortical column establishes its own set of reference frames. We have proposed that cortical columns create reference frames using cells that are equivalent to grid cells and place cells. (Hawkins, 2021, p. 129)

Currently, ML systems have nothing on par with the brain referencing system proposed by Hawkins. An AI may not be able to differentiate between a moon and a yellow traffic light (Atkinson, 2021). This is because it does not know what a moon is and what a traffic light is. Because AI cannot understand these frames of reference, including this, is the criticism of Hinton, the scientist responsible for their development in the 1980s (Hinton, 1981; Zemel, Mozer & Hinton, 1989), they cannot learn the world as a structure.

Thus, it does not seem correct, in the terms of neuroscience (responsible for the study of brain activities), to measure the intelligence of a machine by the results obtained in a singular task. Intelligence, according to Hawkins, which I agree with, is determined by how we grasp and store knowledge about the world. We are intelligent not because we can do one thing particularly well, but because we can learn to do anything.

3.1 Consciousness and Artificial Intelligence

Another aspect that Hawkins' theory elucidates and clarifies is the relationship between consciousness, machines, and human beings. Although this is an avoided theme in neuroscience because of its abstraction and because we don't know what consciousness means, Hawkins still manages to physically explain its aspects.

Consciousness is a process and action (awareness) that can be seen as follows. Hawkins suggests imagining if we could reset our brain to the exact moment we woke up this morning. However, before this reset happened, we would normally do all the things and activities that we were already programmed to do. Then all memories about what we did that day would be erased. Later, after that reset, we would believe that we just woke up. If someone said that we had a snack, breakfast, worked or any other task we would deny it vehemently. Of course, we were conscious while doing these activities. Only the memories were deleted to induce us that we were not. Howe-ver, this demonstrates your sense of consciousness as action (awareness), which many people would call being conscious and requires the moment-by-moment formation of memories about our actions.

> Consciousness also requires that we form moment-to-moment memories of our thoughts. Recall that thinking is just a sequential activation of neurons in the brain. We can remember a sequence of thoughts just as we can remember the sequence of notes in a melody. If we didn't remember our thoughts, we would be unaware of why we were doing anything. For example, we have all experienced going to a room in our house to do something but, upon entering the room, forgetting what we went there for. When this happens, we often ask ourselves, "Where was I just before I got here and what was I thinking?" We try to recall the memory of our recent thoughts so we know why we are now standing in the kit-chen. (Hawkins, 2021, p. 133)

For Hawkins, to feel conscious (awareness) is to feel present, to feel that we are an agent acting in the world, and this is the heart and the main idea of what it means to be conscious. If an artificial intelligence could contemplate models of the world equivalent to the states of the nervous synapses of neurons in brains, and it could remember states and events through memorization, it could be considered conscious, in a way. But this is a step that has not been taken so far. Thus, nothing that we call AI today is intelligent.

No machine has the flexibility and ability to shape itself according to its experiences in the world. However, there is no denying its ability to be built in the near or distant future. One of the main obstacles is precisely understanding what intelligence is, which ends up being left aside by computer scientists.

But we must consider that intelligent machines if properly made, will not be equal to human beings. Intelligence is the ability of the system to learn a model of the world. However, our brain is not only made up of the neocortex but also of the brainstem responsible for emotions and other more primitive behaviors. When truly intelligent machines are created, they should follow the unique arrangements of the neocortex, ignoring the emotional part by choice: since we can choose which parts go in and which parts don't, emotion will possibly be left out due to its complexity. Thus, an AI, des-pite needing the flexibility of human behavior and understanding different models of the world, they do not need the basic instincts of survival or procreation, for example.

The components that Hawkins assumes for the elaboration of an AI, therefore, are the following: embodiment; old brain parts; and the neocortex. Since there is a need for interaction and contact with the world through sensors, embodiment becomes necessary precisely to develop the sense of a world in an AI. However, most ML network does not have an embodiment. There are no sensors attached to the world that can capture inputs and transmit them to the reference frames to develop the world model of AI. Without this embodiment, what can be grasped is limited.

Furthermore, Hawkins assumes the existence of a system equivalent to the brainstem. The purpose would not be to replicate emotions but to move the AI's body. The neocortex does not directly control the movements. It needs a command from the brainstem to act. The same should happen in an AI. The author questions whether it should be this way or whether we could build intelligent machines that where the system equivalent to the neocortex and would directly control the movements. Hawkins does not believe that because the neocortex needs to be connected to something that already has sensors and already has behaviors. In other words, it is not meant to create new behaviors, but to learn how to connect them in new and useful ways.

Hawkins also posits that an AI must possess goals and motivations. Human beings have complex goals and wills. Some are tied to our genes, such as sexual desires, food, and shelter. But it is important to remember that the neocortex itself does not create these goals, motivations, or emotions. It is only actively involved in how these motivations and life goals influence behavior, but it does not lead and determine it. The brainstem is responsible for this task.

And indeed, there is a sense in your statement. Since we are talking about beings that are remarkably intelligent for their respective activities in the social order, to develop machines without any objective or pretension and implement them would be to echo an electronic void without any correspondence. I cannot visualize any sense in an intelligent being that does not act on its own with the means to pursue and build itself.

Finally, for Hawkins AI must have something equivalent to the neocortex. This is the third assumption that a machine must have because it is considered the storage place of intelligence. But learning should not be equated with copying. As every human being is a blank sheet of paper and the world conceptions are built throughout his life through reference frames, the same should happen with an AI. It is not enough that it is inserted into a platform to replicate data and knowledge previously acquired.

Conclusions

We can see computer science's proposals as a utopian choice, and somehow do not match with neuroscience. We must be skeptical, although nothing prevents someone from being anxious about what may come, about the world we currently live in. No human being is endowed with all knowledge. This premise is not reduced by insufficient intelligence, but because no person can be everything and do everything. Some limitations affect human beings, and the same must be true of machines.

In this way, I can make some final considerations. First, Intelligence in its neuroscientific meaning is the modular capacity to learn from the world because learning occurs in the neocortex through the association between neurons, cortical columns, and reference frames. Second, an AI must have the same capabilities as a brain to be called intelligent. Third, the above proposition seems intelligible because the brain is the only reference to what we understand as intelligent. In the current scenario, machines cannot be classified as intelligent because of the gap between neuroscience knowledge and ML development.

I emphasize that the propositions are only final considerations that can serve as a basis for further studies. The purpose of this work was not to bring indisputable parameters but to bring new discursive possibilities about the concept of AI.

References

Atkinson, J. (2021, July 26th). *Tesla's Autopilot Misunderstood the Moon For A Yellow Traffic Light. Video: Automatic.* Swords Today. https://swordstoday.ie/teslas-autopilot-misunderstood-the-moon-for-a-yellow-traffic-light-video-automatic/

Barton, R. A. (1996). Neocortex size and behavioural ecology in primates. *Proceedings of the Royal Society of London*, *263*(1367), 173-177.

Beer, J. M., Fisk, A. D. & Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, *3*(2), 74-99. https://doi.org/10.5898%2FJHRI.3.2.Beer

Binder, J. R. (2017). Current controversies on Wernicke's area and its role in language. *Current neurology and neuroscience reports*, *17*(8), 1-10.

Bostrom, N. & Yudkowsky, E. (2011). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (eds.), *The Cambridge Handbook of Artificial Intelligence* (316-334). Cambridge University Press.

Bostrom, N. (2014). *Superintelligence*. Oxford University Press.

Brayne, S. (2020). *Predict and surveil: Data, discretion, and the future of policing*. Oxford University Press.

Byrne, R. W., & Corp, N. (2004). Neocortex size predicts deception rate in primates. *Proceedings of the Royal Society of London B, 271*(1549), 1693-1699.

Crawford, K. (2021). *The Atlas of AI.* Yale University Press.

Crick, F. H. (1979). Thinking about the brain. *Scientific American*, *241*(3), 219-233.

Davis, R., Shrobe, H. & Szolovits, P. (1993). What is a knowledge representation? AI *magazine*, *14*(1), 17-33. https://doi.org/10.1609/aimag.v14i1.1029

Deaner, R. O., Isler, K., Burkart, J. & Van Schaik, C. (2007). Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain, behavior and evolution*, *70*(2), 115-124.

Edelman, G. M. & Mountcastle, V. B. (1978). T*he mindful brain*: *cortical organization and the group-selective theory of higher brain function*. MIT Press.

Etard, O., Mellet, E., Papathanassiou, D., Benali, K., Houdé, O., Mazoyer, B. & Tzourio-Mazoyer, N. (2000). Picture naming without Broca's and Wernicke's area. N*euroreport*, *11*(3), 617-622.

Felleman, D. J. & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. C*erebral cortex*, *1*(1), 1-47. https://doi.org/10.1093/cercor/1.1.1-a

Franklin, S. & Graesser, A. (1997). Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In J. P. Müller, M. J. Wooldridge & N. R. Jennings (eds), I*ntelligent* A*gents* III A*gent* T*heories*, A*rchitectures*, *and* L*anguages* (pp. 21-35). Springer. https://doi.org/10.1007/BFb0013570

Gharbi, R. B., Elsharkawy, A. M. & Karkoub, M. (1999). Universal neural-network-based model for estimating the PVT properties of crude oil systems. E*nergy & fuels*, *13*(2), 454-458. https://doi.org/10.1021/ef980143v

Haggard, P. (2017). Sense of agency in the human brain. N*ature* R*eviews* N*euroscience*, *18*(4), 196-207.

Hawkins, J. & Ahmad, S. (2016). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. F*rontiers in* N*eural* C*ircuits*, *10*. https://doi.org/10.3389/fncir.2016.00023

Hawkins, J. (2021). A *thousand brains*: A *new theory of intelligence*. Basic Books.

Hinton, G. F. (1981). A parallel computation that assigns canonical object-based frames of reference. In A. Drinan (ed.), P*roceedings of the 7th international joint conference on* A*rtificial intelligence-*V*olume 2* (pp. 683-685). Morgan Kaufmann Publishers Inc.

Huang, H. M., Pavek, K., Ragon, M., Jones, J., Messina, E. & Albus, J. (2007). Characterizing unmanned system autonomy: Contextual autonomous capability and level of autonomy analyses. In G. R. Gerhart, D. W. Gage & C. M. Shoemaker (eds.), U*nmanned* S*ystems* T*echnology* IX. P*roceedings* V*olume* 6561. D*efense and* S*ecurity* S*ymposium | 9-13 April 2007.* International Society for Optics and Photonics.

Kim, D., & Thayer, S. A. (2001). Cannabinoids inhibit the formation of new synapses between hippocampal neurons in culture. J*ournal of* N*euroscience*, *21*(10), RC146. https://doi.org/10.1523/JNEUROSCI.21-10-j0004.2001

Konflanz, D. M. (2019). I*nvestigating hierarchical temporal memory networks applied to dynamic branch prediction* [undergraduate thesis, Universidade Federal da Fronteira Sul]. Digital Repository. https://rd.uffs.edu.br/handle/prefix/3374

Li, W., Ma, L., Yang, G. & Gan, W. B. (2017). REM sleep selectively prunes and maintains new synapses in development and learning. N*ature neuroscience*, *20*(3), 427-437.

Lindenfors, P. (2005). Neocortex evolution in primates: the 'social brain' is for females. B*iology letters*, *1*(4), 407-410.

Luck, M. & d'Inverno, M. (1995, June). A Formal Framework for Agency and Autonomy. In L. Gasser & V. Lesser (eds.), P*roceedings of the* F*irst* I*nternational* C*onference on* M*ultiagent* S*ystems* (pp. 254-260). MIT Press.

Luck, M. & d'Inverno, M. (2001). A conceptual framework for agent definition and development. T*he* C*omputer* J*ournal*, *44*(1), 1-20.

Markman, A. B. (2013). K*nowledge representation*. Psychology Press.

Michaud, A. (2016). Intelligence and Early Mastery of the Reading Skill. *Journal of Biometrics & Biostatistics*, *7*(4). https://doi.org/10.4172/2155-6180.1000327

Naeser, M. A., Helm-Estabrooks, N., Haas, G., Auerbach, S. & Srinivasan, M. (1987). Relationship between lesion extent in Wernicke's area' on computed tomographic scan and predicting recovery of comprehension in Wernicke's aphasia. *Archives of Neurology*, *44*(1), 73-82.

Ogawa, M., Miyata, T., Nakajima, K., Yagyu, K., Seike, M., Ikenaka, K., Yamamoto, H. & Mikoshiba, K. (1995). The *reeler* gene-associated antigen on Cajal-Retzius neurons is a crucial molecule for laminar organization of cortical neurons. *Neuron*, *14*(5), 899-912. https://doi.org/10.1016/0896-6273(95)90329-1

Qiu, Y., Garg, D., Zhou, L., Kharangate, C. R., Kim, S. M. & Mudawar, I. (2020). An artificial neural network model to predict mini/micro-channels saturated flow boiling heat transfer coefficient based on universal consolidated data. *International Journal of Heat and Mass Transfer*, *149*. https://doi.org/10.1016/j.ijheatmasstransfer.2019.119211

Ramón Y Cajal, S. (1923). *Recuerdos de mi vida*. Imprenta de Juan Pueyo.

Gil, V., Nocentini, S. & Del Río, J. A. (2014). Historical first descriptions of Cajal–Retzius cells: from pioneer studies to current knowledge. *Frontiers in neuroanatomy*, *8*, 32, 1-9.

Russell, S. J. & Norvig, P. (2010). *Artificial Intelligence-A Modern Approach* (3. internat. ed.) Pearson Education.

Shrestha, Y. R., Ben-Menahem, S. M. & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, *61*(4), 66-83.

Sun, Q. & Ertekin, T. (2015, April 27th-30th). *The development of artificial-neural-network-based universal proxies to study steam assisted gravity drainage* (SAGD) *and cyclic steam stimulation* (*CSS*) *processes* [paper presented In Society of Petroleum Engineers —SPE— Western Regional Meeting]. Garden Grove, California, USA. https://doi.org/10.2118/174074-MS

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*: *an introduction*. The MIT Press.

Tehovnik, E. J. & Slocum, W. M. (2004). Behavioural state affects saccades elicited electrically from neocortex. *Neuroscience & Biobehavioral Reviews*, *28*(1), 13-25. https://doi.org/10.1016/j.neubiorev.2003.10.001

Yokoi, A. & Diedrichsen, J. (2019). Neural organization of hierarchical motor sequence representations in the human neocortex. *Neuron*, *103*(6), 1178-1190. https://doi.org/10.1016/j.neuron.2019.06.017

Zemel, R. S., Mozer, M. C. & Hinton, G. E. (1989). TRAFFIC: Recognizing objects using hierarchical reference frame transformations. In D. S. Touretzky (ed.), *Advances in neural information processing systems* (pp. 266-273). Morgan Kaufmann Publishers Inc. https://dl.acm.org/doi/10.5555/2969830.2969863