

Information and artificial intelligence

Alfredo Marcos¹

Fecha de recepción: 15/06/2020 - Fecha de aceptación: 23/07/2020

DOI: <https://doi.org/10.22490/26655489.4103>

Resumen

Existe un intenso debate sobre la llamada inteligencia artificial (IA) -alimentada hoy por grandes cantidades de datos (*big data*)- y sobre sus posibles efectos en la vida humana. Me pregunto en este artículo si al esclarecer el concepto de información obtenemos alguna ventaja para pensar la IA. Defenderé que, en efecto, un concepto de información bien elaborado aporta mucha luz sobre lo que llamamos IA. Para argumentar a favor de esta tesis, expondré primero mis ideas sobre la noción de información. Después plantearé muy en breve algunos debates relativos a la IA. Desde el concepto de información podemos entender mejor lo que es un dato, también podemos discutir con más perspectiva la misma existencia de algo que se pueda llamar con propiedad IA, así como las posibilidades que las máquinas tienen de reemplazar al ser humano, de dañar o de mejorar su vida.

Palabras clave: *Inteligencia artificial, megadatos, concepto de información, medida de la información.*

Abstract

There is a deep discussion about the so-called artificial intelligence (AI) –which is currently fuelled by *bigdata*- and about its possible effects on human life. I question in this article if we could get some advantage thinking about AI by clarifying the concept of information. I will argue, in effect, that a well-developed concept of information can provide some light on what we call AI. To argue in favour of this thesis, I will first explain my

¹ Doctor en Filosofía por la Universidad de Barcelona, es profesor de Filosofía de la Ciencia en la Universidad de Valladolid. Orcid: <https://orcid.org/0000-0003-2101-5781>. amarcos@fyl.uva.es

approach to the notion of information. Then, I will raise in brief some deliberations about AI. We can better understand what data is by taking into account the concept of information. We may also discuss, with a higher sense of perspective, even the very existence of something that could properly be called AI, as well as the possibility of machines superseding human beings, to damage or improve their life.

Keywords: *Artificial Intelligence, Big Data, information concept, information measure.*

1. What is information?²

The concept of information appears in a wide range of disciplines, from journalism to biology, for this purpose it is said that we live in the information society. And, of course, it is a key concept in computing, where it often appears quantified. However, the notion of information refers to a group of different concepts and measures, the connection of which is not always clear. To summarize, let's say that information may be seen, and historically has been seen, either as a *thing* (as a substance) or as a *property* of a thing, or as a relation between two or three things. Hereafter, I will argue that the most general concept of information is the one referred as a *triadic relationship* (Peirce, 1931-1935, vol. 5, p. 484).

The information (I), thus understood, will consist of a relation between: *i*) a message (m), which can be linguistic or any other event, *ii*) a reference system (S), about which the message informs the receiver, and *iii*) a receiver (R).

The receiver is a formal scheme residing in a specific subject (e.g. a computer, a robot, a human being or any other living being, an ecosystem ...) Although a specific subject may have more than one receiver and use them alternately. That is, a subject may have different hypotheses or expectations, or change from one to another due to a learning process.

² I have written at length about the concept and measure of information in Marcos, 2011 and 2013. I here summarize the ideas presented in those texts.

It may seem surprising that the issuer, or the source, has not been mentioned. This is because the sender becomes S if the information that R receives, thanks to the message, turns out to be about the issuer itself. On the other hand, in determining the intended meaning, the issuer acts as a virtual receiver, and thus can be formally treated. And finally, there is often no specific sender of the information, especially in non-linguistic contexts.

Most of the conceptual problems related to information occur by ellipsis. Information in a message is often discussed without reference to the receiver or to the reference system. One element of the informational relationship is taken, although the explanation of the other two is forgotten, there is always the implicit suggestion that they exist. The information, if it really is such, goes beyond the message, through a receiver, to a reference system. In other words, the information is not located in the message, nor in the receiver, nor in the reference system, but in the triadic relationship between them.

The relation between the three mentioned elements (m, R, S) will be informative if and only if it produces a change in the expectations that the receiver has about the reference system. From there, we can measure the amount of information based on the magnitude of this change.

We can therefore describe information (I) as a relationship between a message (m), a receiver (R) and a reference system (S). Thus, the higher the probability estimated by the receiver for a given state of the reference system, the more information it will receive when a message indicates that the system is in some other state.

2. How we can measure information

Information can be measured by the effect it produces, that is, by the magnitude of the change that a message produces in the expectations that the receiver has on a certain reference system.

There are two minimum requirements to take a measure of information as correct: it must capture the core of our intuitive notion of information, and it must be consistent with the best

information theory available to us, that is, the mathematical theory of communication of Claude E. Shannon (Shannon & Weaver, 1949). To develop such a measure, we have to specify the concepts advanced here above.

- i. Let's consider a message, m_i , as an element of a set of alternative messages, $M = \{m_1, m_2, \dots, m_n\}$
- ii. S can be any system. $\sigma = \{s_1, s_2, \dots, s_q\}$ is the set of possible states of system S .
- iii. R is characterized by:
 - A set of a priori probabilities associated with the different possible states of the reference system: $P(s_1), \dots, P(s_q)$, where $\sum_k P(s_k) = 1$.
 - and a function that assigns posterior probabilities, $P(s_k|m_i)$, to each pair $\langle m_i, s_k \rangle$; where $\sum_k P(s_k|m_i) = 1$.

The information from- m_i -for- R -about- S can then be measured, taking into account the difference, D , between the a priori probabilities and the posterior probabilities after receiving the message. The a priori probabilities, that is, before receiving the message, are: $P(s_1), \dots, P(s_q)$. The posterior probabilities, that is, after receiving the message, are: $P(s_1|m_i), \dots, P(s_q|m_i)$. The difference is calculated using the formula:

$$D(m_i, R, S) = \sum_k |P(s_k) - P(s_k|m_i)|$$

In accordance with what has been said, the proposed measure of the information will be given by the formula:

$$I(m_i, R, S) = -\log (1-(D/2))$$

It can be proven that:

$$0 \leq D \leq 2$$

If $D = 0$, then $I = 0$

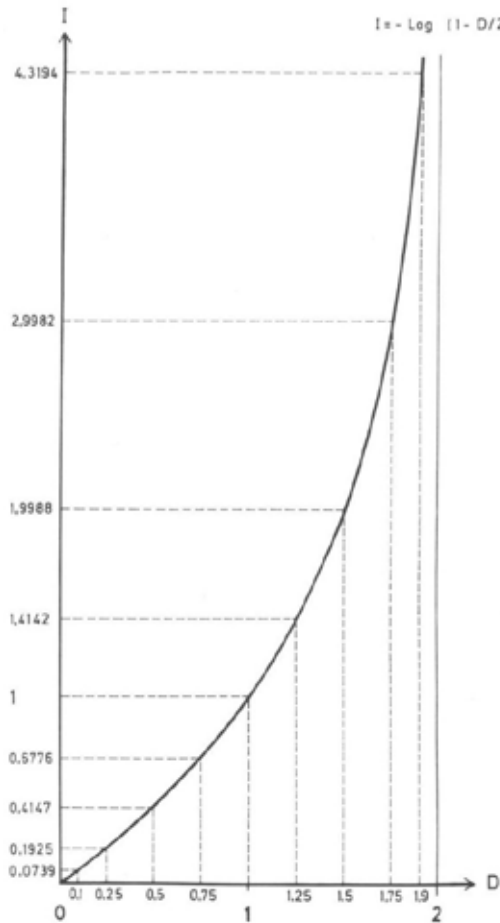
If $0 < D < 2$, then I tends to ∞ as D tends to 2

If $D = 2$, then there is no real value for I

Let's now analyze the meaning of the different values that D and I take:

- i. $D = 0$ means that there is no change in R 's expectations of S , despite receiving the message. In this case it is reasonable that I is equal to Zero. If a message doesn't change my expectations of something, it just doesn't give me information about that something.
- ii. I tends to ∞ if D tends to 2. This means that the greater the clash with the recipient's previous convictions (without reaching the pure contradiction), the greater the information that the receiver has received about the reference system with the arrival of the message (see figure 1).
- iii. $D = 2$ only occurs if the message, m , reports that something that R previously considered impossible is given, that is, the system is in a state whose a priori estimated probability was zero. In this case, I has no real value. This situation can be considered as an indication that a radical restructuring of the subject's expectations is required. The receiver used so far by the subject, its expectations scheme, has collapsed and the passage to a different receiver is required.

Figure 1.- Measurement of the information, I , based on the variation in knowledge, D .



These results are consistent with our intuitive notion of information. Thus, the first of the conditions required to consider a measure of the information is fulfilled. The second is also true, that is, a reinterpretation of Shannon's entropy can be offered in terms of the measure of the information proposed here³.

³ I here avoid demonstration, as it appears in the posts referenced above, and is not required for the rest of my argument.

3. The experience of the impossible ($D=2$)

This last case ($D = 2$) is of great importance, since many learning processes (biological and cultural evolution, Piaget's cognitive development, Kuhn's dynamics of scientific theories...) seem to present two different types of change: cumulative or gradual (assumable within the limits of a certain receptor) and reorganizational or saltational (that requires a radical change, that is, the use of a new receptor). The measure of information I present formally captures this distinction between gradual (in which $0 < D < 2$) and revolutionary (in which $D = 2$) cognitive changes, a distinction that will be crucial in AI discussions, as we will see further below.

According to Kuhn's interpretation of the history of science (2017), we can identify periods in which the new data fit within pre-existing theoretical frameworks or paradigms. In informational terminology, we would say that the new data yields a certain amount of information for a given receiver. There are other periods when traditional expectations are broken, and a paradigm shift is required. In other words, the need for a change of receptor is perceived by the scientific community.

There are many other learning processes that admit the same formal characterization. For example, as Piaget (1970) observed, young children believe that two containers of the same height contain the same amount of liquid. The child uses a receiver in which the height of the liquid is strictly correlated with the total amount of liquid. However, the frustration of this system of expectations will come at one time or another. Then, if the child's intelligence develops normally and within a community, his/her vision of this aspect of the world will change and will move to the use of another receptor. For the new receiver, the total amount of liquid in a container will be correlated with volume and not simply height. It is an abrupt change preceded by a frustrating experience, with all the emotions it refers.

On the other hand, an evolutionary phenomenon such as the extinction of species can also be understood as the emergence of new environmental circumstances that challenge the system of expectations associated with a certain class of organisms. If neither of them is in a position to change receivers, extinction

inevitably comes. If, on the contrary, part of the population, by chance, already had an alternative receptor, then this part can survive

It seems that the learning processes in which human beings intervene are not simply driven by chance, but there is in them a properly free element, voluntary and rational. When interaction with the world fits in with our expectations system, we use it and we learn. Which is rational. But, when we are startled by *the experience of the impossible*, when our expectations system collapses, we can survive thanks to the fact that we can move on to another one. And this step does not have to be purely arbitrary, random or irrational, but, in some sense, is guided by a practical and social knowledge that Aristotle called *phronesis*. This knowledge facilitates the integrative constitution of experience, the management of emotions linked to the frustration of expectations, the propaedeutic of the creative moment and the critical filtering of systems of emerging expectations.

All that has been seen supports the hypothesis that information theory can serve to formally capture the difference between an interaction that fits into previous schemas and one that forces them to break these schemas and create others. In the latter case, the leap towards a new scheme, paradigm, or, in informational terms, receiver, will be driven by a creative element and guided by some form of prudential rationality, if we do not want to trust everything at random.

What does all this tell us about AI and its more advanced processes - deep learning?

4. What we call AI

The term "Artificial Intelligence" is ambiguous. It is used to refer to a field of knowledge and research. In this sense, it is a discipline included within computer science. In fact, this is how the *DRAE (Dictionary of the Royal Academy of the Spanish Language)* defines it: "A scientific discipline that deals with creating computer programs that perform operations comparable to those performed by the human mind, such as learning or logical reasoning". So far the term is clear and unobjectionable. Intelligence, in this case, belongs to the people who do research in this discipline.

But it also serves to refer to the supposed intelligence that some artificial artifacts or systems would have. In this second sense, the term is to be used only under poetic license. That is, strictly speaking, there are no smart artifacts. Intelligence is a natural phenomenon, strange to the artificial, as I will try to argue.

As it is well known, the term AI was coined in 1956, during the already famous Dartmouth Workshop (*Dartmouth Summer Research Project on Artificial Intelligence*), by the American scientist John McCarthy. It is true that, from the rhetorical and propagandistic point of view, this term has been a complete success, since it awakens in the human mind a series of futuristic spells, promises and threats, very suggestive and attractive, almost hypnotic, a mixture of veneration and fear. It is a term that works very well in literature and fictional cinema, in media debates and bestselling essays, as well as in the loquacity aimed at obtaining funds for research.

But, let's think for a moment about what intelligence means. Again we can go to the relevant meanings of the *DRAE*: "Ability to understand or comprehend [...] Ability to solve problems." We know that an artificial system is incapable of understanding or comprehending. It can't even properly be said that a machine counts or computes. Counting involves bringing together two (or more) different moments, and keeping them as such, in a single conscious representation⁴, which a machine does not do. On the other hand, it is true that AI can help us solve problems (computation, geolocation, logistics, telephone assistance, assistance with medical diagnosis, advertising and a long etcetera). But these problems are not for the artificial system, but for the designer or its user.

For a facial recognition system, whether or not to recognize an offender is not a problem. It is a problem for people's safety, and the system can help us deal with it. Of course, the same system can serve to control the population of an imaginary country and to facilitate political repression there. But this is also not a problem for the cameras or for the software involved. It is, undoubtedly, for the suffering subjects of the imaginary country. Neither hammers, nor abacuses, nor the most advanced computer systems have problems. The problems as such are ours, as living beings and as human beings. Only a living being can

4 I address this issue at length in Marcos, 2012.

die or suffer, only people can wonder about the meaning of their lives. Those are problems. And both a hammer and a computer network, each in its own way, can help us deal with them (or make them worse)⁵ but this does not make them smart.

We can also see it if we attack the issue from another angle. The so-called AI is sometimes characterized by its simulation capabilities. Appealing to the classic theme of *The Platters*, we could say that AI is *The Great Pretender*. It simulates functions of human intelligence, it is said. But simulating is not being. Simulating intelligence is not the same as being intelligent. And, furthermore, the notion of function inexorably refers to that of a being for which a given effect is functional. Here, artificial systems also depend on the functionality they may have for humans. Outside the human frame, the lights that go on and off on a screen or the movements of a robot are mere effects, they do not fulfill functions.

The story by Miguel de Unamuno entitled *Mecanópolis*⁶ is very illustrative in this regard. In this dystopia, the protagonist reaches a perfectly mechanized city, but completely devoid of inhabitants. With this, the author introduces in the mechanized city a point of view, that of the human protagonist, which allows him to describe the city in intelligible terms, as something more than a mass of moving materials. The day the protagonist appears in the city, the movements of the machines, simple effects until then, begin to be functional. The human point of view even changes its ontology: a piece of metal that rotates on another, for example, becomes the wheel of a tram.

Because the question, at bottom, is an ontological one. Artifacts, in the Aristotelian tradition, are substances only in an accidental sense, by analogy (Kosman, 1987). Living beings, and especially human beings, that is, people, are substances in their own paradigmatic sense. As it is an ontological difference, the hope (or the threat) of canceling it through technological sophistication is illusory, a mere categorical mistake.

⁵ I do not address here the question of possible neutrality of technique. In fact, I think there are good arguments to deny it, but this is not the issue in the present context.

⁶ This story has recently been commented from a philosophical point of view by Alicia Villar and Mario Ramos (2019).

Strictly speaking, as Luc Julia (2019), one of the creators of the AI system named SIRI, states, "artificial intelligence does not exist". But, how has the claim on the existence of intelligent artificial systems been reached? There was a time when the design of a symbolic AI was intended, using explicit rules, restricted thanks to certain heuristic strategies. On the other hand, the Connectionist Route was attempted. This methodology consists of the modeling of neural networks, which are inspired by the functioning of the networks of authentic neurons that populate our nervous system. From the 1960s there was an explosion of AI research, first in the symbolist line and then, from the 1970s, on the connectionist side. At this time, expectations grew enormously. But in the late 1980s, AI research began to give symptoms of exhaustion. The connectionist methods, which seemed more promising, were based on statistical learning, which requires a large computing capacity and a huge amount of data, both superior to what was available at the time.

However, with the rise of personal computers, the miniaturization and lower cost of components, and the arrival of the Internet, both computing capacity and the contribution of data skyrocketed in a few years. Some companies soon saw in this takeoff a promise of abundance, and their own activity generated even more data and greater capacity to process it. Google, Amazon, Facebook, Apple and Microsoft (collectively known as GAFAM), as well as Chinese Baidu, Alibaba, Tencent, Xiaomi (BATX) and Huawei, have rekindled the illusion for AI. They have done so on the basis of brute force, that is, at the cost of *big data* and an exponentially growing and increasingly affordable computing capacity.

Thanks to this rebound in AI research, very useful and precise systems have been achieved in various fields. It is always about systems specialized in a certain function. Undoubtedly, machine translation systems, facial or visual recognition systems, expert systems in medicine or finance, automatic driving systems or robots that simulate conversational skills (*chatbots*) have improved. None of them understand what they do, but that does not diminish its functionality... for us. No strong or general AI system has been achieved. That is, an artificial system that could cope with all the functions that human intelligence performs in an integrated way.

Information and artificial intelligence

Alfredo Marcos¹

Fecha de recepción: 15/06/2020 - Fecha de aceptación: 23/07/2020

DOI: <https://doi.org/10.22490/26655489.4103>

Resumen

Existe un intenso debate sobre la llamada inteligencia artificial (IA) -alimentada hoy por grandes cantidades de datos (*big data*)- y sobre sus posibles efectos en la vida humana. Me pregunto en este artículo si al esclarecer el concepto de información obtenemos alguna ventaja para pensar la IA. Defenderé que, en efecto, un concepto de información bien elaborado aporta mucha luz sobre lo que llamamos IA. Para argumentar a favor de esta tesis, expondré primero mis ideas sobre la noción de información. Después plantearé muy en breve algunos debates relativos a la IA. Desde el concepto de información podemos entender mejor lo que es un dato, también podemos discutir con más perspectiva la misma existencia de algo que se pueda llamar con propiedad IA, así como las posibilidades que las máquinas tienen de reemplazar al ser humano, de dañar o de mejorar su vida.

Palabras clave: *Inteligencia artificial, megadatos, concepto de información, medida de la información.*

Abstract

There is a deep discussion about the so-called artificial intelligence (AI) –which is currently fuelled by *bigdata*- and about its possible effects on human life. I question in this article if we could get some advantage thinking about AI by clarifying the concept of information. I will argue, in effect, that a well-developed concept of information can provide some light on what we call AI. To argue in favour of this thesis, I will first explain my

¹ Doctor en Filosofía por la Universidad de Barcelona, es profesor de Filosofía de la Ciencia en la Universidad de Valladolid. Orcid: <https://orcid.org/0000-0003-2101-5781>. amarcos@fyl.uva.es

approach to the notion of information. Then, I will raise in brief some deliberations about AI. We can better understand what data is by taking into account the concept of information. We may also discuss, with a higher sense of perspective, even the very existence of something that could properly be called AI, as well as the possibility of machines superseding human beings, to damage or improve their life.

Keywords: *Artificial Intelligence, Big Data, information concept, information measure.*

1. What is information?²

The concept of information appears in a wide range of disciplines, from journalism to biology, for this purpose it is said that we live in the information society. And, of course, it is a key concept in computing, where it often appears quantified. However, the notion of information refers to a group of different concepts and measures, the connection of which is not always clear. To summarize, let's say that information may be seen, and historically has been seen, either as a *thing* (as a substance) or as a *property* of a thing, or as a relation between two or three things. Hereafter, I will argue that the most general concept of information is the one referred as a *triadic relationship* (Peirce, 1931-1935, vol. 5, p. 484).

The information (I), thus understood, will consist of a relation between: *i*) a message (m), which can be linguistic or any other event, *ii*) a reference system (S), about which the message informs the receiver, and *iii*) a receiver (R).

The receiver is a formal scheme residing in a specific subject (e.g. a computer, a robot, a human being or any other living being, an ecosystem ...) Although a specific subject may have more than one receiver and use them alternately. That is, a subject may have different hypotheses or expectations, or change from one to another due to a learning process.

² I have written at length about the concept and measure of information in Marcos, 2011 and 2013. I here summarize the ideas presented in those texts.

It may seem surprising that the issuer, or the source, has not been mentioned. This is because the sender becomes S if the information that R receives, thanks to the message, turns out to be about the issuer itself. On the other hand, in determining the intended meaning, the issuer acts as a virtual receiver, and thus can be formally treated. And finally, there is often no specific sender of the information, especially in non-linguistic contexts.

Most of the conceptual problems related to information occur by ellipsis. Information in a message is often discussed without reference to the receiver or to the reference system. One element of the informational relationship is taken, although the explanation of the other two is forgotten, there is always the implicit suggestion that they exist. The information, if it really is such, goes beyond the message, through a receiver, to a reference system. In other words, the information is not located in the message, nor in the receiver, nor in the reference system, but in the triadic relationship between them.

The relation between the three mentioned elements (m, R, S) will be informative if and only if it produces a change in the expectations that the receiver has about the reference system. From there, we can measure the amount of information based on the magnitude of this change.

We can therefore describe information (I) as a relationship between a message (m), a receiver (R) and a reference system (S). Thus, the higher the probability estimated by the receiver for a given state of the reference system, the more information it will receive when a message indicates that the system is in some other state.

2. How we can measure information

Information can be measured by the effect it produces, that is, by the magnitude of the change that a message produces in the expectations that the receiver has on a certain reference system.

There are two minimum requirements to take a measure of information as correct: it must capture the core of our intuitive notion of information, and it must be consistent with the best

information theory available to us, that is, the mathematical theory of communication of Claude E. Shannon (Shannon & Weaver, 1949). To develop such a measure, we have to specify the concepts advanced here above.

- i. Let's consider a message, m_i , as an element of a set of alternative messages, $M = \{m_1, m_2, \dots, m_n\}$
- ii. S can be any system. $\sigma = \{s_1, s_2, \dots, s_q\}$ is the set of possible states of system S .
- iii. R is characterized by:
 - A set of a priori probabilities associated with the different possible states of the reference system: $P(s_1), \dots, P(s_q)$, where $\sum_k P(s_k) = 1$.
 - and a function that assigns posterior probabilities, $P(s_k|m_i)$, to each pair $\langle m_i, s_k \rangle$; where $\sum_k P(s_k|m_i) = 1$.

The information from- m_i -for- R -about- S can then be measured, taking into account the difference, D , between the a priori probabilities and the posterior probabilities after receiving the message. The a priori probabilities, that is, before receiving the message, are: $P(s_1), \dots, P(s_q)$. The posterior probabilities, that is, after receiving the message, are: $P(s_1|m_i), \dots, P(s_q|m_i)$. The difference is calculated using the formula:

$$D(m_i, R, S) = \sum_k |P(s_k) - P(s_k|m_i)|$$

In accordance with what has been said, the proposed measure of the information will be given by the formula:

$$I(m_i, R, S) = -\log (1-(D/2))$$

It can be proven that:

$$0 \leq D \leq 2$$

If $D = 0$, then $I = 0$

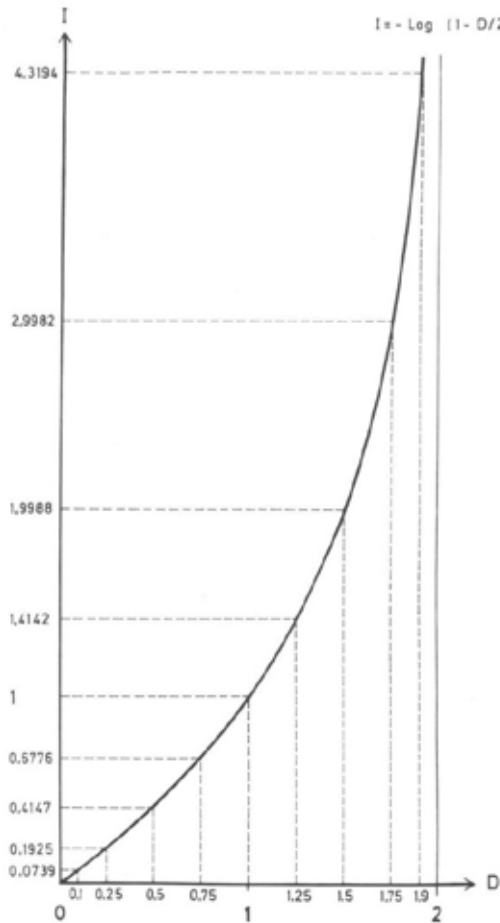
If $0 < D < 2$, then I tends to ∞ as D tends to 2

If $D = 2$, then there is no real value for I

Let's now analyze the meaning of the different values that D and I take:

- i. $D = 0$ means that there is no change in R 's expectations of S , despite receiving the message. In this case it is reasonable that I is equal to Zero. If a message doesn't change my expectations of something, it just doesn't give me information about that something.
- ii. I tends to ∞ if D tends to 2. This means that the greater the clash with the recipient's previous convictions (without reaching the pure contradiction), the greater the information that the receiver has received about the reference system with the arrival of the message (see figure 1).
- iii. $D = 2$ only occurs if the message, m , reports that something that R previously considered impossible is given, that is, the system is in a state whose a priori estimated probability was zero. In this case, I has no real value. This situation can be considered as an indication that a radical restructuring of the subject's expectations is required. The receiver used so far by the subject, its expectations scheme, has collapsed and the passage to a different receiver is required.

Figure 1.- Measurement of the information, I, based on the variation in knowledge, D.



These results are consistent with our intuitive notion of information. Thus, the first of the conditions required to consider a measure of the information is fulfilled. The second is also true, that is, a reinterpretation of Shannon's entropy can be offered in terms of the measure of the information proposed here³.

³ I here avoid demonstration, as it appears in the posts referenced above, and is not required for the rest of my argument.

3. The experience of the impossible ($D=2$)

This last case ($D = 2$) is of great importance, since many learning processes (biological and cultural evolution, Piaget's cognitive development, Kuhn's dynamics of scientific theories...) seem to present two different types of change: cumulative or gradual (assumable within the limits of a certain receptor) and reorganizational or saltational (that requires a radical change, that is, the use of a new receptor). The measure of information I present formally captures this distinction between gradual (in which $0 < D < 2$) and revolutionary (in which $D = 2$) cognitive changes, a distinction that will be crucial in AI discussions, as we will see further below.

According to Kuhn's interpretation of the history of science (2017), we can identify periods in which the new data fit within pre-existing theoretical frameworks or paradigms. In informational terminology, we would say that the new data yields a certain amount of information for a given receiver. There are other periods when traditional expectations are broken, and a paradigm shift is required. In other words, the need for a change of receptor is perceived by the scientific community.

There are many other learning processes that admit the same formal characterization. For example, as Piaget (1970) observed, young children believe that two containers of the same height contain the same amount of liquid. The child uses a receiver in which the height of the liquid is strictly correlated with the total amount of liquid. However, the frustration of this system of expectations will come at one time or another. Then, if the child's intelligence develops normally and within a community, his/her vision of this aspect of the world will change and will move to the use of another receptor. For the new receiver, the total amount of liquid in a container will be correlated with volume and not simply height. It is an abrupt change preceded by a frustrating experience, with all the emotions it refers.

On the other hand, an evolutionary phenomenon such as the extinction of species can also be understood as the emergence of new environmental circumstances that challenge the system of expectations associated with a certain class of organisms. If neither of them is in a position to change receivers, extinction

inevitably comes. If, on the contrary, part of the population, by chance, already had an alternative receptor, then this part can survive

It seems that the learning processes in which human beings intervene are not simply driven by chance, but there is in them a properly free element, voluntary and rational. When interaction with the world fits in with our expectations system, we use it and we learn. Which is rational. But, when we are startled by *the experience of the impossible*, when our expectations system collapses, we can survive thanks to the fact that we can move on to another one. And this step does not have to be purely arbitrary, random or irrational, but, in some sense, is guided by a practical and social knowledge that Aristotle called *phronesis*. This knowledge facilitates the integrative constitution of experience, the management of emotions linked to the frustration of expectations, the propaedeutic of the creative moment and the critical filtering of systems of emerging expectations.

All that has been seen supports the hypothesis that information theory can serve to formally capture the difference between an interaction that fits into previous schemas and one that forces them to break these schemas and create others. In the latter case, the leap towards a new scheme, paradigm, or, in informational terms, receiver, will be driven by a creative element and guided by some form of prudential rationality, if we do not want to trust everything at random.

What does all this tell us about AI and its more advanced processes - deep learning?

4. What we call AI

The term "Artificial Intelligence" is ambiguous. It is used to refer to a field of knowledge and research. In this sense, it is a discipline included within computer science. In fact, this is how the *DRAE (Dictionary of the Royal Academy of the Spanish Language)* defines it: "A scientific discipline that deals with creating computer programs that perform operations comparable to those performed by the human mind, such as learning or logical reasoning". So far the term is clear and unobjectionable. Intelligence, in this case, belongs to the people who do research in this discipline.

But it also serves to refer to the supposed intelligence that some artificial artifacts or systems would have. In this second sense, the term is to be used only under poetic license. That is, strictly speaking, there are no smart artifacts. Intelligence is a natural phenomenon, strange to the artificial, as I will try to argue.

As it is well known, the term AI was coined in 1956, during the already famous Dartmouth Workshop (*Dartmouth Summer Research Project on Artificial Intelligence*), by the American scientist John McCarthy. It is true that, from the rhetorical and propagandistic point of view, this term has been a complete success, since it awakens in the human mind a series of futuristic spells, promises and threats, very suggestive and attractive, almost hypnotic, a mixture of veneration and fear. It is a term that works very well in literature and fictional cinema, in media debates and bestselling essays, as well as in the loquacity aimed at obtaining funds for research.

But, let's think for a moment about what intelligence means. Again we can go to the relevant meanings of the *DRAE*: "Ability to understand or comprehend [...] Ability to solve problems." We know that an artificial system is incapable of understanding or comprehending. It can't even properly be said that a machine counts or computes. Counting involves bringing together two (or more) different moments, and keeping them as such, in a single conscious representation⁴, which a machine does not do. On the other hand, it is true that AI can help us solve problems (computation, geolocation, logistics, telephone assistance, assistance with medical diagnosis, advertising and a long etcetera). But these problems are not for the artificial system, but for the designer or its user.

For a facial recognition system, whether or not to recognize an offender is not a problem. It is a problem for people's safety, and the system can help us deal with it. Of course, the same system can serve to control the population of an imaginary country and to facilitate political repression there. But this is also not a problem for the cameras or for the software involved. It is, undoubtedly, for the suffering subjects of the imaginary country. Neither hammers, nor abacuses, nor the most advanced computer systems have problems. The problems as such are ours, as living beings and as human beings. Only a living being can

⁴ I address this issue at length in Marcos, 2012.

die or suffer, only people can wonder about the meaning of their lives. Those are problems. And both a hammer and a computer network, each in its own way, can help us deal with them (or make them worse)⁵ but this does not make them smart.

We can also see it if we attack the issue from another angle. The so-called AI is sometimes characterized by its simulation capabilities. Appealing to the classic theme of *The Platters*, we could say that AI is *The Great Pretender*. It simulates functions of human intelligence, it is said. But simulating is not being. Simulating intelligence is not the same as being intelligent. And, furthermore, the notion of function inexorably refers to that of a being for which a given effect is functional. Here, artificial systems also depend on the functionality they may have for humans. Outside the human frame, the lights that go on and off on a screen or the movements of a robot are mere effects, they do not fulfill functions.

The story by Miguel de Unamuno entitled *Mecanópolis*⁶ is very illustrative in this regard. In this dystopia, the protagonist reaches a perfectly mechanized city, but completely devoid of inhabitants. With this, the author introduces in the mechanized city a point of view, that of the human protagonist, which allows him to describe the city in intelligible terms, as something more than a mass of moving materials. The day the protagonist appears in the city, the movements of the machines, simple effects until then, begin to be functional. The human point of view even changes its ontology: a piece of metal that rotates on another, for example, becomes the wheel of a tram.

Because the question, at bottom, is an ontological one. Artifacts, in the Aristotelian tradition, are substances only in an accidental sense, by analogy (Kosman, 1987). Living beings, and especially human beings, that is, people, are substances in their own paradigmatic sense. As it is an ontological difference, the hope (or the threat) of canceling it through technological sophistication is illusory, a mere categorical mistake.

⁵ I do not address here the question of possible neutrality of technique. In fact, I think there are good arguments to deny it, but this is not the issue in the present context.

⁶ This story has recently been commented from a philosophical point of view by Alicia Villar and Mario Ramos (2019).

Strictly speaking, as Luc Julia (2019), one of the creators of the AI system named SIRI, states, "artificial intelligence does not exist". But, how has the claim on the existence of intelligent artificial systems been reached? There was a time when the design of a symbolic AI was intended, using explicit rules, restricted thanks to certain heuristic strategies. On the other hand, the Connectionist Route was attempted. This methodology consists of the modeling of neural networks, which are inspired by the functioning of the networks of authentic neurons that populate our nervous system. From the 1960s there was an explosion of AI research, first in the symbolist line and then, from the 1970s, on the connectionist side. At this time, expectations grew enormously. But in the late 1980s, AI research began to give symptoms of exhaustion. The connectionist methods, which seemed more promising, were based on statistical learning, which requires a large computing capacity and a huge amount of data, both superior to what was available at the time.

However, with the rise of personal computers, the miniaturization and lower cost of components, and the arrival of the Internet, both computing capacity and the contribution of data skyrocketed in a few years. Some companies soon saw in this takeoff a promise of abundance, and their own activity generated even more data and greater capacity to process it. Google, Amazon, Facebook, Apple and Microsoft (collectively known as GAFAM), as well as Chinese Baidu, Alibaba, Tencent, Xiaomi (BATX) and Huawei, have rekindled the illusion for AI. They have done so on the basis of brute force, that is, at the cost of *big data* and an exponentially growing and increasingly affordable computing capacity.

Thanks to this rebound in AI research, very useful and precise systems have been achieved in various fields. It is always about systems specialized in a certain function. Undoubtedly, machine translation systems, facial or visual recognition systems, expert systems in medicine or finance, automatic driving systems or robots that simulate conversational skills (*chatbots*) have improved. None of them understand what they do, but that does not diminish its functionality... for us. No strong or general AI system has been achieved. That is, an artificial system that could cope with all the functions that human intelligence performs in an integrated way.

Let's think, to get a more concrete idea, of a useful AI system, for example, for banking. It helps us decide whether or not we should grant a certain loan. The bank wants to grant it to whoever is going to return it and deny it to the alleged defaulters. How do we know, before an applicant, whether or not he/she will be a good payer? A couple of data may be taken, such as the age and salary of the person concerned. In other words, we represent the applicant by a point on a coordinate diagram. Age is on one axis and salary on another. We go to the bank's historical archive. We represent clients, those who have paid and those who have failed, by points on our diagram. We fit a Gaussian Process Regression, so that above it the good payers predominate and below are the defaulters. We see in which sector is the point that represented our applicant. If it is above, loan granted and denied otherwise. This conventional system is very imprecise. It produces false positives and false negatives. But thanks to AI we can improve it considerably, even making it very precise. For this we use many more types of data (age, salary, average balance in the last year, in the last five, type of employment...). And of course, we exponentially increase the number of data of each type. Not only for the applicant, but also in the long-standing record. We can add all the types of data that we want, since they already exist in abundance. In fact, all of us lavishly gifted them to the GAFAM and BATX on duty. Let's put in the long-standing record, then, the hobbies of customers, their consumption habits, whether or not they are gamblers, whether or not they buy books, if they read this or that digital, what trips they make and in which vehicles, what their blood group is, or the sequence of their genome... and so on as far as we want.

With such a rich history, we will trace a space not of two, but of many dimensions. And we will not resign ourselves to a poor regression process, but the AI system, after a not very long period of training, will offer us a probably irregular distribution, but very precise in an n -dimensional space. Once we locate in this space the point that represents our applicant, we will see if he/she is surrounded by compliers or defaulters and based on this we will make a decision.

Or the AI system will make it for us. Is the human financial adviser superfluous? Is the doctor's opinion superfluous when an AI system diagnoses cancer? Does the human pilot strike when the intelligent navigation system decides to vary the speed or

Let's think, to get a more concrete idea, of a useful AI system, for example, for banking. It helps us decide whether or not we should grant a certain loan. The bank wants to grant it to whoever is going to return it and deny it to the alleged defaulters. How do we know, before an applicant, whether or not he/she will be a good payer? A couple of data may be taken, such as the age and salary of the person concerned. In other words, we represent the applicant by a point on a coordinate diagram. Age is on one axis and salary on another. We go to the bank's historical archive. We represent clients, those who have paid and those who have failed, by points on our diagram. We fit a Gaussian Process Regression, so that above it the good payers predominate and below are the defaulters. We see in which sector is the point that represented our applicant. If it is above, loan granted and denied otherwise. This conventional system is very imprecise. It produces false positives and false negatives. But thanks to AI we can improve it considerably, even making it very precise. For this we use many more types of data (age, salary, average balance in the last year, in the last five, type of employment...). And of course, we exponentially increase the number of data of each type. Not only for the applicant, but also in the long-standing record. We can add all the types of data that we want, since they already exist in abundance. In fact, all of us lavishly gifted them to the GAFAM and BATX on duty. Let's put in the long-standing record, then, the hobbies of customers, their consumption habits, whether or not they are gamblers, whether or not they buy books, if they read this or that digital, what trips they make and in which vehicles, what their blood group is, or the sequence of their genome... and so on as far as we want.

With such a rich history, we will trace a space not of two, but of many dimensions. And we will not resign ourselves to a poor regression process, but the AI system, after a not very long period of training, will offer us a probably irregular distribution, but very precise in an n -dimensional space. Once we locate in this space the point that represents our applicant, we will see if he/she is surrounded by compliers or defaulters and based on this we will make a decision.

Or the AI system will make it for us. Is the human financial adviser superfluous? Is the doctor's opinion superfluous when an AI system diagnoses cancer? Does the human pilot strike when the intelligent navigation system decides to vary the speed or

height of flight? Would acting against the output of an AI system always be reckless? What or who is responsible in each case for a possible failure? Should we put quotation marks in “decide” and “diagnose”? It is not necessary to draw much attention to the important consequences of all kinds that these questions have. We immediately glimpse its philosophical, anthropological, social, legal, economic, moral, labor implications ... Of course, I do not aspire here to answer each of the questions, or even record their numerous consequences. However, perhaps the informational approach can shed light on all of this.

5. An informational interpretation of AI

Let's start with the notion of data. Perhaps we could understand it in terms of what information theory calls a message. In fact, any event becomes *data* only if it is *given* to a receiver. An ink stain is not a piece of information, unless someone can read it as an accounting entry, for example. Nor is data the electromagnetic or quantum state of a piece of matter, unless that piece of matter is what is called a component, that is, unless it is integrated into a system that can take the state in question as a piece of data. There is no data that is not integrated into a more complex system, which implies spaces of possibilities and a receiver that connects them to a reference system. The ink stain has a certain shape, but it could have another shape within a certain space of possibilities that the reader knows. The same is true of the electromagnetic states of a component of a computer system. These states are *data-for*, not simply data. They are data because they are for the system as a whole.

But the same could be said of the computer system as a whole. Said system is now in a very complex electromagnetic state, which only turns out to be a datum, and not a simple fact in the world, insofar as this state can be taken by a human being as a datum. Otherwise: the concept of data, like that of a message, implies that of a receiver of those data, and said receiver, directly or indirectly, will be a conscious human being. It is my gaze on my computer which transforms into data for my computer all the *inputs* that come to it, and into data for me all the *outputs* that it emits.

Only by assuming a consciousness in the background, a kind of transcendental self, can we speak of certain states of matter as data, and of an AI system as a receiver of data. This receiver can be formally characterized in the way we have done above with the information receivers. An AI system is a system of expectations (or rather a prosthesis of our systems of expectations). It places a point in an n -dimensional space built from a data history, and, based on that, it tells us what to expect from the object represented by that point. Like any expectations system, it can collapse when it records the occurrence of something that it previously considered impossible. It is what we have called *the experience of the impossible here above* ($D = 2$). When this occurs, the receiver itself is left without adaptability, it cannot learn from this experience.

An AI system aimed at granting credits - to follow the example - can assume that some of its predictions fail, that some of the loans granted fail, and can integrate these new data into the history, learn from it and reorganize its geometry. What it cannot assume is the sudden collapse, the abrupt failure, of all the credits in force, even the most solvent ones. If this happens, it is not the AI system that should react, but the managers of the bank in question. And they will react, first of all, drastically changing expectations. They can do so given that they are not artificial systems, but conscious people who can understand the phenomenon, something that is not expected from the machine, and who can start their creativity to generate better expectations systems from now on.

Let's think that the collapse of all the credits could have been due to a seismic or climatic or astronomical phenomenon, but also due to a cultural trend, a political movement or an epidemic ... A person can come to *understand* what happened. It is the person, the human being who has general intelligence, who can replace the collapsed receiver with another. Only one person can intentionally connect and reconnect the logical (digital) layout with the physical (analog) layout.

In logical terms, we can build expectations by induction, generalizing past correlations. This is done very well by people and the rest of living beings, even from very few cases. Recent AI systems are also excellent at this type of task, although they require massive contributions of historical data. Then, expectations

systems can be applied to new cases and thus guide our action. Many times we do it deductively. But when expectations fail, we have to try to understand what is going on. We do this by creating hypotheses that, if correct, would explain the situation. This type of inference is called abduction. Abduction is creative, it forces us to get out of given expectation systems. And here people are essential, with their peculiar ontology, with their conscience, intentionality and interaction with the world. It was Charles S. Peirce who first studied in depth these kinds of essential inferences to confront the new elements, the unexpected, or previously considered impossible, the extraordinary. And it was the Spanish poet Jorge Guillén who wrote in sententious verse: "The extraordinary: everything."

6. Conclusion

The so-called AI works well as a prosthesis for human intelligence. It is intelligence in the same way that a prosthetic hand is a hand. It takes on entity and meaning only in the framework of human action. The first reason, and the most obvious one, is that it originated as a product of human action. But there are those who think that once started it could self-maintain, even self-improve beyond the human limits. However, the problem regarding the limits of AI is not technical, it is ontological.

The truth is that there would not even be data properly (or big data) in the absence of a human receptor, of a person who is capable of unifying in consciousness the dispersion of the gross facts, who is capable of giving intentionality to each of them, thus connecting some parts of reality, which play as data, with others about which these data tell us something. In addition, however lax we may be in the selection of data for the long-standing, there will always be some selection, since we cannot feed a machine with all the data. It is true that today we play with huge amounts of data. In this sense, and perhaps for the first time, we swim in abundance. But *much* data never becomes *all* data. In collecting them and in their contribution to AI systems, we continually make inexorably human judgments of relevance.

Nor does AI make sense as a *great pretender*, that is, as a simulator of intellectual functions, unless we know how to distinguish

between functions and simple effects. This is a distinction that leads us inexorably to the existence of a living being and, in many cases, of a person. The electromagnetic states of a system only simulate functions if there are functions to simulate. And the success in a cancer diagnosis, or in the granting of a credit or in a piloting action, is only functional for people and for the communities in which they live.

On the other hand, only people can make decisions. The very concept of decision is foreign to the mechanical. What we call "decision" in an AI system will be so only to the extent that a human being has made the genuine decision to delegate some action to the system, that is, to automate it. The ultimate responsibility, whether things turn out right or wrong, can only be from a human being.

And it is our responsibility to rely on the best AI systems available when making decisions. That is, in many cases, it will be convenient for the doctor to rely on an AI system to support the diagnosis. Especially if the system is well calibrated and has been shown to be useful and reliable in clinical trials. Acting against the indications of the system would burden the doctor with an extra responsibility, while following these indications would somewhat alleviate his responsibility in case of error. But, in any case, the human expert must have his hands free to act even against the indications of an AI system, since it may be the case - and it happens - that said system collapses before what we have called the experience of impossible, given what we have characterized in informational terms as $D = 2$. Here you need *someone* who understands, who creatively searches for new explanations, who reconsiders and generates new systems of expectations.

Summarizing, and in relation to the so-called AI, we could say that what is intelligent is not artificial and what is artificial is not intelligent.

Bibliography

- David, Marie & Sauviat, Cédric (2019). *Intelligence Artificielle*. Mónaco: Éditions du Rocher.
- Julia, Luc (2019). *L'intelligence artificielle n'existe pas*. París: Éditions First.
- Kosman, L. A. (1987). "Animals and Other Beings in Aristotle", en Gotthelf, A. y Lennox, J. (eds.). *Philosophical Issues in Aristotle's Biology*. Cambridge: Cambridge University Press.
- Kuhn, Thomas (2017). *La estructura de las revoluciones científicas*. Ciudad de México: FCE, 4ª edición.
- Marcos, Alfredo (2011). "Bioinformation as a Triadic Relation", en G. Terzis & R. Arp (eds.), *Information and Living Systems*. Cambridge, MA: M.I.T. Press, pp. 55-90.
- Marcos, Alfredo & Arp, Robert (2013). "Information in the Biological Sciences", en K. Kampourakis (ed.), *The Philosophy of Biology. A Companion for Educators*. Dordrecht: Springer, pp. 511-548.
- Marcos, Alfredo (2012). "La relación de semejanza como principio de inteligibilidad de la naturaleza", en F. Rodríguez Valls (ed.), *La inteligencia en la naturaleza*. Madrid: Biblioteca Nueva, pp. 73-94.
- Peirce, Chales S. (1931-1935). *Collected Papers*. Cambridge, MA: Harvard University Press.
- Piaget, Jean (1970). *L'Épistemologie gènétique*. París: P.U.F..
- Shannon, Claude & Weaver, Warren (1949). *Mathematical theory of communication*. Urbana: University of Illinois Press.
- Villar, Alicia & Ramos, Mario (2019). "Mecanópolis: una distopía de Miguel de Unamuno". *Pensamiento*, 75 (283): 321-343.

