

INFORMACIÓN E INTELIGENCIA ARTIFICIAL

INFORMATION AND ARTIFICIAL INTELLIGENCE

ALFREDO MARCOS

Universidad de Valladolid (España)
<http://orcid.org/0000-0003-2101-5781>
amarcos@fyl.uva.es

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 an intense debate about the so-called artificial intelligence (AI), which is currently fueled by *big data*, and about its possible effects on human life. I wonder in this article if by clarifying the concept of information we could get some advantage in thinking about AI. I will argue that, in effect, a well-developed concept of information can shed some of light on what we call AI. To argue in favour of this thesis, I will first explain my approach to the notion of information. Then, I will raise in brief some debates about AI. We can better understand what a data is by taking into account the concept of information. We can also discuss with more perspective even the very existence of something that can be properly called AI, as well as the possibilities that machines have to replace the human beings, to damage or to improve their life.

Keywords: artificial Intelligence, big data, information concept, information measure.

1. ¿Qué es la información?

El concepto de información¹ aparece en las más diversas disciplinas, desde el periodismo a la biología. No en vano se dice que vivimos en la sociedad de la información. Y, claro está, es un concepto clave en informática, donde incluso con frecuencia aparece cuantificado. Sin embargo, la noción de información refiere a una familia de conceptos y medidas diferentes, cuya conexión no siempre es clara. Por resumir, digamos que la información puede ser vista, e históricamente ha sido vista, o bien como una *cosa* (como una sustancia), o bien como una *propiedad* de una cosa, o bien como una *relación* entre dos o entre tres cosas. En lo sucesivo defenderé que el concepto más general de información es el que la trata como una *relación triádica* (Peirce, 1931-1935, vol. 5, p. 484).

La información (I), así entendida, consistirá en una relación entre: *i*) un mensaje (m), que puede ser cualquier evento, lingüístico o de otro tipo, *ii*) un sistema de referencia (S), acerca del cual el mensaje informa al receptor, y *iii*) un receptor (R).

El receptor es un esquema formal residente en un sujeto concreto (un computador, un robot, un ser humano u otro ser vivo, un ecosistema...). Si bien un sujeto concreto puede disponer de más de un receptor y usarlos alternativamente. Es decir, puede jugar con distintas hipótesis o expectativas, o cambiar de una a otra debido a un proceso de aprendizaje.

Pudiera parecer sorprendente que el emisor, o la fuente, no haya sido mencionado. Ello es debido a que el emisor se convierte en S si la información que R recibe, gracias al mensaje, resulta ser sobre el propio emisor. Por otro lado, al determinar el significado intentado, el emisor actúa como un receptor virtual, y así puede ser tratado formalmente. Y, por último, a menudo no existe un emisor determinado de la información, especialmente en contextos no lingüísticos.

La mayor parte de los problemas conceptuales relativos a la información ocurren por elipsis. A menudo se habla de la información de un mensaje sin hacer ninguna referencia al receptor ni al sistema de referencia. Se toma un elemento de la relación informacional, pero se olvida explicitar los otros dos, aunque siempre queda la sugerencia implícita de que existen. La información, si de veras es tal, va más allá del mensaje, a través de un receptor, hasta un sistema de referencia. Dicho de otro modo, la información no está ubicada en el mensaje, ni en receptor, ni en el sistema de referencia, sino en la relación triádica ente ellos.

La relación entre los tres elementos mencionados (m, R, S) será informativa si y solo si produce un cambio en las expectativas que el receptor tiene sobre el sistema de referencia. A partir de ahí, podremos medir la cantidad de información en función de la magnitud de dicho cambio.

Podemos, pues, describir la información (I) como una relación entre un mensaje (m), un receptor (R) y un sistema de referencia (S). Así, cuanto más alta sea la probabilidad estimada por el receptor para un determinado estado del sistema de referencia, más información recibirá cuando un mensaje le indique que el sistema está en otro estado.

2. Cómo podemos medir la información

La información puede ser medida por el efecto que produce, es decir, por la magnitud del cambio que un mensaje produce en las expectativas que el receptor tiene sobre un cierto sistema de referencia.

Dos son los requisitos mínimos para dar por correcta una medida de la información: debe captar el núcleo de nuestra noción intuitiva de información y debe ser coherente con la mejor teoría de la información de

¹ He escrito por extenso sobre el concepto y la medida de la información en Marcos (2011, 2013). Resumo aquí las ideas expuestas en estos textos.

que disponemos, es decir, la teoría matemática de la comunicación de Claude E. Shannon (Shannon & Weaver, 1949). Para desarrollar una medida así, tenemos que precisar los conceptos adelantados más arriba.

- i) Consideremos un mensaje, m_i , como un elemento de un conjunto de mensajes alternativos, $M = \{m_1, m_2, \dots, m_n\}$
- ii) S puede ser cualquier sistema. $\sigma = \{s_1, s_2, \dots, s_q\}$ es el conjunto de estados posibles del sistema S.
- iii) R viene caracterizado por:
 - iii.i) Un conjunto de probabilidades a priori asociadas con los diferentes estados posibles del sistema de referencia: $P(s_1), \dots, P(s_q)$, siendo $\sum_k P(s_k) = 1$.
 - iii.ii) y una función que asigna probabilidades a posteriori, $P(s_k | m_i)$, a cada par $\langle m_i, s_k \rangle$; siendo $\sum_k P(s_k | m_i) = 1$.

La información de- m_i -para-R-sobre-S, puede ser medida, entonces, tomando en cuenta la diferencia, D, entre las probabilidades a priori y las probabilidades a posteriori de la recepción del mensaje. Las probabilidades a priori, o sea, antes de la recepción del mensaje, son: $P(s_1), \dots, P(s_q)$. Las probabilidades a posteriori, es decir, tras la recepción del mensaje, son: $P(s_1 | m_i), \dots, P(s_q | m_i)$. La diferencia se calcula mediante la fórmula:

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

Conforme a lo dicho, la medida propuesta de la información vendrá dada por la fórmula:

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

Se puede probar que:

$$0 \leq D \leq 2$$

Si $D = 0$, entonces $I = 0$

Si $0 < D < 2$, entonces I tiende a ∞ según D tiende a 2

Si $D = 2$, entonces no hay valor real para I

Analícemos ahora el significado de los diferentes valores que toman D e I:

i) $D = 0$ significa que no hay cambio en las expectativas que R tiene sobre S, a pesar de la recepción del mensaje. En este caso es razonable que I sea igual a cero. Si un mensaje no cambia mis expectativas respecto de algo, simplemente no me aporta información sobre ese algo.

ii) I tiende a ∞ si D tiende a 2. Esto significa que cuanto mayor es el choque con las convicciones previas del receptor (sin llegar a la pura contradicción), mayor es la información que el receptor ha recibido sobre el sistema de referencia con la llegada del mensaje (véase figura 1).

iii) $D = 2$ sólo se produce si el mensaje, m, informa de que se da algo que R consideraba previamente como imposible, es decir, de que el sistema está en un estado cuya probabilidad estimada a priori era cero. En este caso, I no tiene un valor real. Esta situación puede ser considerada como una indicación de que se requiere una reestructuración radical de las expectativas del sujeto. El receptor usado hasta el momento por el sujeto, su esquema de expectativas, ha colapsado y se requiere el paso a un receptor distinto.

Estos resultados son coherentes con nuestra noción intuitiva de información. Se cumple así la primera de las condiciones exigidas para dar por buena una medida de la información. También se cumple la segunda, es decir, se puede ofrecer una reinterpretación de la entropía de Shannon en términos de la medida de la información aquí propuesta.²

² Evito aquí la demostración, pues figura en las publicaciones referidas más arriba, y no se requiere para el resto de mi argumento.

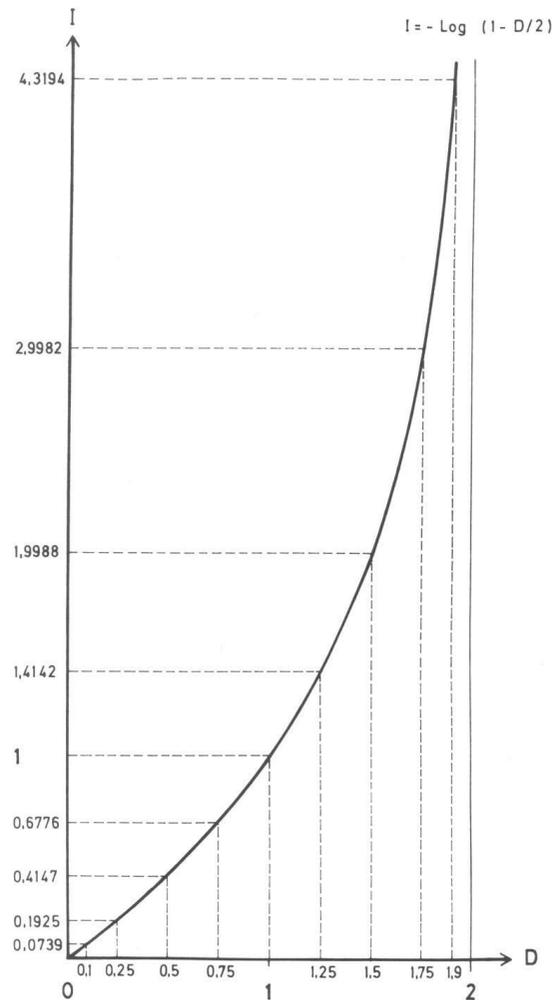


Figura 1. Medida de la información I en función de la variación del conocimiento D .

3. La experiencia de lo imposible ($D=2$)

Este último caso ($D=2$) tiene gran importancia, ya que muchos procesos de aprendizaje (evolución biológica y cultural, desarrollo piagetiano de estructuras cognitivas, dinámica kuhniana de las teorías científicas...) parece que presentan dos tipos diferentes de cambio: acumulativo o gradual (asumible dentro de los límites de un determinado receptor) y reorganizativo o saltacional (que requiere un cambio radical, o sea, la utilización de un nuevo receptor). La medida de la información que presento capta formalmente esta distinción entre cambios cognoscitivos graduales (en los cuales $0 < D < 2$) y revolucionarios (en los cuales $D=2$), distinción que será crucial en los debates sobre IA, como veremos más abajo.

De acuerdo con la interpretación que hace Kuhn (2017) de la historia de la ciencia, podemos identificar periodos en los que los nuevos datos se acomodan dentro de marcos teóricos o paradigmas preexistentes. En terminología informacional, diríamos que los nuevos datos rinden una cierta cantidad de información para un receptor dado. Existen otros periodos en los que las expectativas tradicionales resultan rotas y se requiere un cambio de paradigma. Es decir, la necesidad de un cambio de receptor es sentida por la comunidad científica.

Existen otros muchos procesos de aprendizaje que admiten la misma caracterización formal. Por ejemplo, como observó Piaget (1970), los niños pequeños creen que dos recipientes de la misma altura contienen

la misma cantidad de líquido. El niño usa un receptor en el que la altura del líquido está estrictamente correlacionada con la cantidad total de líquido. No obstante, la frustración de este sistema de expectativas llegará en un momento u otro. Entonces, si la inteligencia del niño se desarrolla con normalidad y en el seno de una comunidad, cambiará su visión de este aspecto del mundo y pasará a la utilización de otro receptor. Para el nuevo receptor, la cantidad total de líquido en un recipiente estará correlacionada con el volumen y no simplemente con la altura. Es un cambio abrupto precedido de una experiencia de frustración, con todo su cortejo de emociones.

Por otro lado, un fenómeno evolutivo como la extinción de especies también puede ser entendido como el surgimiento de nuevas circunstancias medioambientales que desafían el sistema de expectativas asociadas con una determinada clase de organismos. Si ninguno de ellos está en condiciones de cambiar de receptor, la extinción llega inevitablemente. Si, por el contrario, parte de la población, por azar, disponía ya de un receptor alternativo, entonces, esta parte puede sobrevivir

Da la impresión de que los procesos de aprendizaje en los que intervienen seres humanos no están pilotados simplemente por el azar, sino que hay en ellos un elemento propiamente libre, voluntario y racional. Cuando la interacción con el mundo encaja con nuestro sistema de expectativas, empleamos este y vamos aprendiendo. Lo cual es racional. Pero, cuando nos sobresalta *la experiencia de lo imposible*, cuando nuestro sistema de expectativas colapsa, podemos sobrevivir gracias a que podemos pasar a otro. Y este paso no tiene por qué ser puramente arbitrario, azaroso o irracional, sino que, en algún sentido, está guiado por un saber práctico y social que Aristóteles llamó *phrónesis*. Dicho saber nos facilita la constitución integradora de la experiencia, la gestión de las emociones vinculadas a la frustración de expectativas, la propedéutica del momento creativo y el filtrado crítico de los sistemas de expectativas emergentes.

Todo lo visto apoya la hipótesis de que la teoría de la información puede servir para captar formalmente la diferencia entre una interacción que encaja en esquemas previos y una que obliga a romper dichos esquemas y a crear otros. En este último caso, el salto hacia un nuevo esquema, paradigma, o, en términos informacionales, receptor, estará impulsado por un elemento creativo y guiado por alguna forma de racionalidad prudencial, si es que no queremos fiarlo todo al azar.

¿Qué nos dice todo ello sobre la IA y sus procesos de aprendizaje más avanzados (*deep learning*)?

4. A qué llamamos IA

El término “inteligencia artificial” es ambiguo. Sirve para referirnos a un campo de conocimiento e investigación. En este sentido, es una disciplina incluida dentro de las ciencias de la computación. De hecho, así lo define el *DRAE*: “Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico”. Hasta aquí el término es claro e inobjetable. La inteligencia, en este caso, pertenece a las personas que investigan en esta disciplina.

Pero también sirve para referirnos a la supuesta inteligencia que tendrían algunos artefactos o sistemas artificiales. En este segundo sentido, el término ha de usarse solo bajo licencia poética. Es decir, estrictamente hablando, no hay artefactos inteligentes. La inteligencia es un fenómeno natural, ajeno a lo artificial, como trataré de argumentar.

Como es bien sabido, el término IA fue acuñado en 1956, durante la ya famosa conferencia de Dartmouth, por el científico estadounidense John McCarthy. Es cierto que, desde el punto de vista retórico y propagandístico, este término ha resultado todo un éxito, pues despierta en la mente humana una serie de embelesos futuristas, de promesas y de amenazas, muy sugerentes y atractivos, casi hipnóticos, una mezcla de veneración y miedo. Es un término que funciona muy bien en la literatura y en el cine de ficción, en los debates mediáticos y en los ensayos con vocación de *bestseller*, así como en la facundia encaminada a la obtención de fondos para la investigación.

Pero, pensemos por un momento qué significa inteligencia. De nuevo podemos acudir a las acepciones pertinentes del *DRAE*: “Capacidad de entender o comprender [...] Capacidad de resolver problemas”. Sabemos que un sistema artificial es incapaz de entender o de comprender. Ni siquiera se puede decir con propiedad que una máquina cuente o compute. Contar implica reunir dos (o más) momentos diferentes, y mantenerlos como tales, en una sola representación consciente (véase Marcos, 2012), cosa que una máquina no hace. Sí es cierto, en cambio, que la IA puede ayudarnos a resolver problemas (cómputo, geolocalización, logística, asistencia telefónica, asistencia al diagnóstico médico, a la publicidad y un largo etcétera). Pero estos problemas no lo son para el sistema artificial, sino para el diseñador o para el usuario del mismo. Para un sistema de reconocimiento facial, el reconocer o no a un delincuente no supone un problema. Es un problema para la seguridad de las personas, y el sistema puede ayudarnos a afrontarlo. Por supuesto, el mismo sistema puede servir para controlar a la población de un imaginario país y para facilitar allí la represión política. Pero esto tampoco es un problema para las cámaras o para el software implicado. Lo es, indudablemente, para los sufridos súbditos del imaginario país. Ni los martillos, ni los ábacos, ni los sistemas informáticos más avanzados tienen problemas. Los problemas los tenemos nosotros, como seres vivos y como seres humanos. Solo un ser vivo puede morir o sufrir, solo una persona puede preguntarse por el sentido de su vida. Esos son problemas. Y tanto un martillo como una red informática, cada uno a su modo, pueden ayudarnos a afrontarlos (o a empeorarlos).³ Pero esto no los convierte en inteligentes.

Podemos verlo también si atacamos la cuestión desde otro ángulo. A veces se caracteriza la llamada IA por su capacidad de simulación. Apelando al tema clásico de *The Platters*, podríamos decir que la IA es *The Great Pretender*. Simula funciones propias de la inteligencia humana, se dice. Pero simular no es ser. Simular la inteligencia no es lo mismo que ser inteligente. Y, además, la noción de función remite inexorablemente a la de un ser para la cual un efecto dado es funcional. Aquí, los sistemas artificiales dependen también de la funcionalidad que puedan tener para el ser humano. Fuera del marco humano, las luces que se encienden y apagan en una pantalla o los movimientos de un robot son meros efectos, no cumplen funciones.

Es muy ilustrativo al respecto el relato de Miguel de Unamuno titulado *Mecanópolis*.⁴ En esta distopía, el protagonista alcanza una ciudad perfectamente mecanizada, pero carente por completo de habitantes. Con ello, el autor introduce en la ciudad mecanizada un punto de vista, el del protagonista humano, lo cual le permite describir la urbe en términos inteligibles, como algo más que un amasijo de materiales en movimiento. El día en que el protagonista aparece por la ciudad, los movimientos de las máquinas, simples efectos hasta entonces, comienzan a ser funcionales. El punto de vista humano cambia incluso su ontología: un pedazo de metal que gira sobre otro, por ejemplo, pasa a ser la rueda de un tranvía.

Porque la cuestión, en el fondo, es de carácter ontológico. Los artefactos, en la tradición aristotélica, son sustancias solo en un sentido accidental, por analogía (Kosman, 1987). Los seres vivos, y en especial los seres humanos, es decir, las personas, sí son sustancias en sentido propio y paradigmático. Al tratarse de una diferencia ontológica, la esperanza (o la amenaza) de anularla por la vía de la sofisticación tecnológica resulta ilusoria, un mero error categorial.

Estrictamente hablando, como afirma Luc Julia (2019), uno de los creadores del sistema de IA denominado SIRI, “la inteligencia artificial no existe”. Pero, ¿cómo se ha llegado a la *pretensión* de que existen sistemas artificiales inteligentes? Hubo un tiempo en que se intentó diseñar una IA simbólica, mediante reglas explícitas, restringidas gracias a ciertas estrategias heurísticas. Por otro lado, se intentó la vía conexionista. Esta metodología consiste en la modelización de redes llamadas neuronales, que se inspiran en el funcionamiento de las redes de auténticas neuronas que pueblan nuestro sistema nervioso. A partir de los años 60 se vivió una explosión de investigaciones en IA, primero en la línea simbolista y después, a partir de los 70, en la vertiente conexionista. En este momento crecieron enormemente las expectativas. Pero a finales de los 80, la investigación en IA empezó a dar síntomas de agotamiento. Los métodos conexionistas, que parecían más prometedores, se basaban en el aprendizaje estadístico, lo cual requiere una gran capacidad de computación y una ingente cantidad de datos, ambas superiores a lo que estaba disponible en la época.

³ No abordo aquí la cuestión de la posible neutralidad de la técnica. De hecho, creo que hay buenos argumentos para negarla, pero no es esta la cuestión en el presente contexto.

⁴ Este relato ha sido recientemente comentado desde un punto de vista filosófico por Villar y Ramos (2019).

No obstante, con el auge de los ordenadores personales, la miniaturización y abaratamiento de los componentes y la llegada de Internet, se dispararon en pocos años tanto la capacidad de computación, como el aporte de datos. Algunas empresas vieron pronto en este despegue una promesa de abundancia, y su propia actividad generó aun muchos más datos y mayor capacidad para procesarlos. Google, Amazon, Facebook, Apple y Microsoft (conocidas en conjunto como GAFAM), así como las chinas Baidu, Alibaba, Tencent, Xiaomi (BATX) y Huawei, han encendido de nuevo la ilusión por la IA. Lo han hecho a base de fuerza bruta, es decir, a costa del *big data* y de una capacidad de computación exponencialmente creciente y cada día más asequible.

Gracias a este repunte de la investigación en IA se han logrado sistemas muy útiles y precisos en diversos ámbitos. Siempre se trata de sistemas especializados en una cierta función. Han mejorado, indudablemente, los sistemas de traducción automática, los de reconocimiento facial o visual, los sistemas expertos en medicina o en finanzas, los sistemas de conducción automática o los robots que simulan habilidades conversacionales (*chatbots*). Ninguno de ellos entiende lo que hace, pero eso no disminuye su funcionalidad... para nosotros. No se ha logrado ningún sistema de IA fuerte o general. Es decir, un sistema artificial que pudiera afrontar de modo integrado todas las funciones que realiza la inteligencia humana.

Pensemos, para hacernos una idea más concreta, en un sistema de IA útil, por ejemplo, para la banca. Nos ayuda a decidir si debemos o no conceder un determinado crédito. El banco quiere concedérselo a quien vaya a devolverlo y denegárselo a los presuntos morosos. ¿Cómo sabemos, ante un solicitante, si va a resultar o no un buen pagador? Se pueden tomar un par de datos, como por ejemplo la edad y el salario del interesado. Dicho de otro modo, representamos al solicitante mediante un punto en un diagrama de coordenadas. En un eje va la edad y en otro el salario. Acudimos al archivo histórico del banco. Representamos a los clientes, a los que han pagado y a los fallidos, mediante puntos en nuestro diagrama. Ajustamos una curva de regresión gaussiana, de modo que por encima de la misma predominen los buenos pagadores y por debajo los morosos. Vemos en que sector queda el punto que representaba a nuestro solicitante. Si está por encima, préstamo concedido. Denegado en caso contrario. Este sistema convencional es muy impreciso. Produce falsos positivos y falsos negativos. Pero gracias a la IA podemos mejorarlo considerablemente, hasta hacerlo muy preciso. Para ello empleamos muchos más tipos de datos (edad, salario, saldo medio en el último año, en los cinco últimos, tipo de empleo...). Y, por supuesto, aumentamos exponencialmente el número de datos de cada tipo. No solo para el solicitante, sino también en el histórico. Podemos añadir todos los tipos de datos que queramos, pues existen ya en abundancia. De hecho todos nosotros los regalamos con prodigalidad a los GAFAM y BATX de turno. Pongamos en el histórico, pues, las aficiones de los clientes, sus hábitos de consumo, si son o no apostadores, si compran o no libros, si leen este o aquel digital, qué viajes hacen y en qué vehículos, cuál es su grupo sanguíneo o la secuencia de su genoma... y así hasta donde queramos.

Con un histórico tan nutrido, trazaremos un espacio no de dos, sino de n dimensiones. Y no nos resignaremos a una pobre curva de regresión, sino que el sistema de IA, tras un periodo no muy largo de entrenamiento, nos ofrecerá una distribución probablemente irregular, pero muy precisa en un espacio n -dimensional. Una vez que ubiquemos en este espacio el punto que representa a nuestro solicitante, veremos si está rodeado de cumplidores o de morosos y en función de ello tomaremos decisión.

O la tomará el sistema de IA por nosotros. ¿Está de más, entonces, el asesor financiero humano?, ¿sobra la opinión del médico cuando un sistema de IA diagnostica un cáncer?, ¿huelga el piloto humano cuando el sistema de navegación inteligente decide variar la velocidad o la altura de vuelo?, ¿podría calificarse siempre de temerario el obrar contra el *output* de un sistema de IA?, ¿qué o quién se hace responsable en cada caso de un posible fallo?, ¿habría que poner comillas en “decide” y en “diagnostica”? No hace falta llamar mucho la atención sobre las importantes consecuencias de todo tipo que tienen estas preguntas.⁵ Vismos inmediatamente sus implicaciones filosóficas, antropológicas, sociales, jurídicas, económicas, morales, laborales... Desde luego, no aspiro aquí a dar respuesta a cada una de las cuestiones, ni a registrar siquiera sus numerosas consecuencias. No obstante, quizá el enfoque informacional pueda arrojar un adarme de luz sobre todo ello.

⁵ Sobre las implicaciones sociopolíticas de la IA puede verse David y Sauviat (2019).

5. Una interpretación informacional de la IA

Comencemos por la noción de dato. Quizá podríamos entenderla en términos de lo que la teoría de la información llama mensaje. De hecho, un suceso cualquiera se convierte en un *dato* solo si es *dado* a un receptor. Una mancha de tinta no es un dato, salvo que alguien pueda leerla como una anotación contable, por ejemplo. Tampoco es un dato el estado electromagnético o cuántico de un pedazo de materia, salvo que ese pedazo de materia sea lo que se llama un componente, es decir, salvo que esté integrado en un sistema que pueda tomar el estado en cuestión como un dato. No hay dato que no esté integrado en un sistema más complejo, lo cual implica unos espacios de posibilidades y un receptor que los conecte con un sistema de referencia. La tinta tiene una cierta forma, pero podría tener otra dentro de un cierto espacio de posibilidades que el lector conoce. Lo mismo sucede con los estados electromagnéticos de un componente de un sistema de computación. Dichos estados son *datos-para*, no simplemente datos. Son datos porque lo son para el sistema en su conjunto.

Pero lo mismo podríamos decir del sistema de computación en su conjunto. Dicho sistema está ahora en un estado electromagnético muy complejo, que solo resulta ser un dato, y no un simple hecho en el mundo, en la medida en que dicho estado puede ser tomado por un ser humano como un dato. De otro modo: el concepto de dato, como el de mensaje, implica el de un receptor de ese dato, y dicho receptor, directa o indirectamente, será un ser humano consciente. Es mi mirada sobre mi ordenador la que transforma en datos para él todos los *inputs* que le llegan, y en datos para mí los *outputs* que emite.

Solo suponiendo una conciencia en el trasfondo, una especie de yo trascendental, podemos hablar de ciertos estados de la materia como datos, y de un sistema de IA como un receptor de datos. Dicho receptor se puede caracterizar formalmente a la manera en que lo hemos hecho más arriba con los receptores de información. Un sistema de IA es un sistema de expectativas (o más bien una prótesis de nuestros sistemas de expectativas). Coloca un punto en un espacio n -dimensional construido a partir de un histórico de datos, y, en función de ello, nos dice qué se puede esperar del objeto representado por ese punto. Como cualquier sistema de expectativas, puede colapsar cuando registra la ocurrencia de algo que previamente consideraba imposible. Es lo que más arriba hemos llamado *la experiencia de lo imposible* ($D=2$). Cuando esto ocurre, el propio receptor se queda sin capacidad de adaptación, no puede aprender de esta experiencia.

Un sistema de IA orientado a la concesión de créditos –por seguir con el ejemplo– puede asumir que algunas de sus predicciones fallen, que alguno de los préstamos concedidos resulten fallidos, y puede integrar estos nuevos datos en el histórico, aprender de ello y reorganizar su geometría. Lo que no puede asumir es el colapso repentino, el fallo abrupto, de todos los créditos en vigor, aun de los más solventes. Si esto sucede, no es el sistema de IA el que debe reaccionar, sino los responsables del banco en cuestión. Y reaccionarán, en primer lugar, cambiando drásticamente las expectativas. Lo pueden hacer dado que no son sistemas artificiales, sino personas conscientes que pueden llegar a *entender* el fenómeno, cosa que no se espera de la máquina, y que pueden poner en marcha su creatividad para generar en adelante mejores sistemas de expectativas.

Pensemos que el colapso de todos los créditos podría haberse debido a un fenómeno sísmico o climático o astronómico, pero también a una moda cultural, a un movimiento político o a una epidemia... Una persona puede llegar a *entender* lo que sucedió. Es la persona, el ser humano que dispone de una inteligencia general, quien puede reemplazar el receptor colapsado por otro. Solo una persona puede conectar y reconectar intencionalmente el plano lógico (digital) con el físico (analógico).

En términos lógicos, podemos construir expectativas por inducción, generalizando correlaciones pasadas. Esto lo hacen muy bien las personas y el resto de los seres vivos, incluso a partir de muy pocos casos. También los sistemas recientes de IA son excelentes en este tipo de tareas, aunque requieren masivos aportes de datos históricos. Después, los sistemas de expectativas se pueden aplicar a nuevos casos y así orientan nuestra acción. Muchas veces lo hacemos de modo deductivo. Pero cuando fallan las expectativas, tenemos que intentar entender lo que está pasando. Lo hacemos creando hipótesis que, de ser correctas, explicarían la situación. Este tipo de inferencia se denomina abducción. La abducción es creativa, nos obliga a salir de los sistemas de expectativas dados. Y aquí las personas son imprescindibles, con su peculiar ontología, con su

conciencia, intencionalidad e interacción con el mundo. Fue Charles S. Peirce quien primero estudió a fondo este tipo de inferencias imprescindibles para enfrentarnos a lo nuevo, a lo inesperado, o considerado previamente como imposible, a lo extraordinario. Y fue el poeta castellano Jorge Guillén quien escribió en verso sentencioso: “Lo extraordinario: todo”.

6. Conclusión

La llamada IA funciona bien como prótesis de la inteligencia humana. Es inteligencia en la misma medida en que una mano protésica es una mano. Cobra entidad y sentido solo en el marco de la acción humana. La primera razón, y la más obvia, es que se originó como producto de la acción humana. Pero hay quien piensa que una vez puesta en marcha podría auto-mantenerse, incluso auto-mejorarse al margen ya de lo humano. No obstante, el problema en cuanto a los límites de la IA no es técnico, es ontológico.

Lo cierto es que ni siquiera existirían propiamente datos (ni *big data*) en ausencia de un receptor humano, de una persona que sea capaz de unificar en la conciencia la dispersión de los hechos brutos, que sea capaz de darle intencionalidad a cada uno de ellos, de conectar así unas partes de la realidad, que juegan como datos, con otras acerca de las cuales nos dicen algo estos datos. Además, por más laxos que podamos ser en la selección de datos para el histórico, siempre habrá alguna selección, pues no podemos alimentar una máquina con *todos* los datos. Es verdad que hoy jugamos con cantidades ingentes de datos. En este sentido, y quizá por primera vez, nadamos en la abundancia. Pero *muchos* datos no llegan a ser nunca *todos* los datos. En la recogida de los mismos y en su aportación a los sistemas de IA realizamos continuamente juicios de relevancia inexorablemente humanos.

Tampoco tiene sentido la IA como *great pretender*, es decir, como simulador de funciones intelectuales, a no ser que sepamos distinguir entre funciones y simples efectos. Es esta una distinción que nos lleva inexorablemente a la existencia de un ser vivo y, en muchos casos, de una persona. Los estados electromagnéticos de un sistema solo simulan funciones si hay funciones que simular. Y el acierto en un diagnóstico de cáncer, o bien en la concesión de un crédito o en una acción de pilotaje, solo es funcional para las personas y para las comunidades en las que estas viven.

Por otro lado, solo las personas pueden tomar decisiones. El mismo concepto de decisión es ajeno a lo mecánico. Lo que llamamos “decisión” en un sistema de IA lo será solo en la medida en que un ser humano haya tomado la genuina decisión de delegar alguna acción en el sistema, es decir, de automatizarla. La responsabilidad última, salgan bien o mal las cosas, solo puede ser de un ser humano. Y es responsabilidad nuestra apoyarnos en los mejores sistemas de IA disponibles a la hora de tomar decisiones. Es decir, en muchos casos, será conveniente que el médico se fie de un sistema de IA para el apoyo al diagnóstico. Especialmente si el sistema está bien calibrado y se ha mostrado útil y fiable en ensayos clínicos. Obrar en contra de las indicaciones del sistema haría cargar al médico con una responsabilidad extra, mientras que seguir dichas indicaciones aliviaría algo su responsabilidad en caso de error. Pero, de cualquier modo, el experto humano debe tener las manos libres para obrar incluso contra las indicaciones de un sistema de IA, pues puede darse el caso –y se da– en que dicho sistema colapse ante lo que hemos llamado la experiencia de lo imposible, ante lo que hemos caracterizado en términos informacionales como $D=2$. Aquí hace falta *alguien* que entienda, que busque creativamente nuevas explicaciones, que replantee y genere nuevos sistemas de expectativas.

Resumiendo, y en relación a la llamada IA, podríamos decir que lo que tiene de inteligente no es artificial y lo que tiene de artificial no es inteligente.

Referencias

- David, Marie y Cédric Sauviat (2019) *Intelligence Artificielle*. Mónaco: Éditions du Rocher.
- Julia, Luc (2019) *L'intelligence artificielle n'existe pas*. París: Éditions First.
- Kosman, Aryeh L. (1987) "Animals and Other Beings in Aristotle", en A. Gotthelf y J. Lennox (eds.), *Philosophical Issues in Aristotle's Biology*. Cambridge: Cambridge University Press.
- Kuhn, Thomas S. (2017) *La estructura de las revoluciones científicas*. Ciudad de México: FCE, 4ª edición.
- Marcos, Alfredo (2011) "Bioinformation as a Triadic Relation", en G. Terzis y R. Arp (eds.), *Information and Living Systems*. Cambridge, MA: MIT Press, pp. 55-90.
- Marcos, Alfredo y Robert Arp (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, Charles S. (1931-1935) *Collected Papers*. Cambridge, MA: Harvard University Press.
- Piaget, Jean (1970) *L'Épistémologie génétique*. París: Presses universitaires de France.
- Shannon, Claude & Warren Weaver (1949) *Mathematical theory of communication*. Urbana, IL: University of Illinois Press.
- Villar, Alicia y Mario Ramos (2019) "Mecanópolis: una distopía de Miguel de Unamuno". *Pensamiento*, 75 (283): 321-343.