

INTELIGENCIA ARTIFICIAL

Reflexiones desde la filosofía y la ética

Diego Gracia

Sumario

El tema de la “inteligencia artificial” ha cobrado en estos últimos años una actualidad no solo sorprendente sino también preocupante. De hecho, ha puesto en guardia a todo tipo de instituciones, que han comenzado a plantearse cómo controlar este sorprendente fenómeno, preocupadas por llegar tarde a su necesaria regulación y control.

Hasta hace relativamente poco tiempo, el debate sobre la inteligencia humana se hacía siempre por contraste con la llamada “inteligencia animal”, de tal modo que los protagonistas eran por lo general etólogos o psicólogos. Hoy el escenario es algo distinto, porque el contraste se establece con la llamada “inteligencia artificial”, debido a la importancia que en las últimas décadas ha cobrado la computación y esa nueva ciencia llamada informática. Ni que decir tiene que esto exige aclarar qué significa inteligencia o de qué estamos hablando cuando usamos esa palabra.

¿Qué es inteligencia? La definición más elemental la identifica con la capacidad de procesar información, algo que hacen tanto los seres vivos como ahora también los ordenadores. Lo que no podían hacer éstos era aprender, algo que parecía privativo de los seres vivos. Pero a día de hoy las máquinas también aprenden, en sistemas como el Chat GPT.

¿Hay algo que diferencie la inteligencia humana de las máquinas y los animales? La filosofía ha respondido por lo general afirmativamente. Y a la pregunta por el origen de esa característica propia y peculiar de la especie humana ha dado dos respuestas, la clásica, que afirma la existencia de una sustancia espiritual en el interior del cuerpo humano, y la emergentista, para la que la novedad del psiquismo humano emerge por pura complejidad de los elementos estructurales del cuerpo. En este caso, se plantea la posibilidad de que la complejidad cada vez mayor de los sistemas informáticos acabe haciendo posible el que en algún momento pueda surgir esa cualidad nueva e irreductible a todo lo anterior que llamamos inteligencia específicamente humana. Algo que, en cualquier caso, parece muy alejado de las capacidades científicas y técnicas actuales.

Introducción

El tema de la “inteligencia artificial” ha cobrado en estos últimos años una actualidad no solo sorprendente sino también preocupante. De hecho, ha puesto en guardia a todo tipo de instituciones, que han comenzado a plantearse cómo controlar este nuevo fenómeno, preocupados por no llegar tarde a su necesaria regulación y control.¹

De la sorpresa inicial se ha ido pasando a la preocupación por el futuro, que se muestra tan incierto como amenazante. De ahí que parezca llegado el momento de iniciar la reflexión sobre este sorprendente fenómeno, que a no dudar tendrá presencia recurrente en esta Academia a partir de ahora. De hecho, Adela Cortina se ocupaba ya de este problema en su reciente artículo *¿Personas electrónicas?*²

De dónde venimos: la inteligencia natural

El término “inteligencia artificial” se ha acuñado sobre un patrón de vieja progenie en nuestra cultura, el de “inteligencia natural”. Los filósofos griegos, y más en concreto Aristóteles, definieron al ser humano, precisamente, por su inteligencia. *Zôon lógon ékhon*, animal dotado de *lógos* lo llama Aristóteles al comienzo de su *Política*.³ En esto ha visto toda la tradición occidental la definición de ser humano, a diferencia de los animales, carentes de *lógos*, dado que no están dotados de lo que Aristóteles llama en su tratado *De anima*, *noûs*, inteligencia o intelecto, una cualidad por demás peculiar, ya que, como afirma, “es separable, sin mezcla e impasible, siendo como es acto por su propia entidad.” Y añade que el *noûs*, “una vez separado, es solo aquello que en realidad es y únicamente esto es inmortal y eterno”.⁴ Tales afirmaciones sobre el intelecto las hace Aristóteles no sin confesar sus dudas al respecto. “Por lo que hace al intelecto y a la potencia especulativa no está nada claro el asunto, si bien parece tratarse de un género distinto de alma y que solamente él puede darse separado como lo eterno de lo corruptible.”⁵ Y más adelante aclara: “La facultad sensible no se da sin el cuerpo, mientras que el intelecto es separable”.⁶ Esta entidad inmortal y eterna, que

¹ Cf. European Commission, *White Paper on Artificial Intelligence*, COM(2020) 65 final.

² A. Cortina, “¿Personas electrónicas?”, *El País* el 7 de septiembre de 2023.

³ *Pol I* 2: 1253 a 9-10.

⁴ *De an III* 5: 429 b 18-23.

⁵ *De an II* 2: 413 b 25-28.

⁶ *De an III* 4: 429 b 5-6.

diferencia a los seres humanos de los animales, tiene la capacidad de abstraer las formas inteligibles de la materia. Los textos citados de Aristóteles son complejos y han dado lugar a diversas interpretaciones. En cualquier caso, la teoría que se impuso fue la de que, a diferencia de las almas de las plantas y los animales, que son materiales, el alma de los seres humanos es espiritual y capaz de penetrar en las formas de las cosas, que son universales, y por tanto captar su esencia.

El debate sobre la sede de la inteligencia: cardiocentrismo y cerebrocentrismo

Afirmado lo anterior, inmediatamente surge la pregunta de dónde asienta ese *noûs* que es tan esencial y definitorio del ser humano. Y aquí las respuestas se dividieron. Hubo un amplio conjunto de pensadores, entre ellos Aristóteles, que fueron “cardiocéntricos”, porque defendieron que la sede del psiquismo superior estaba en el corazón, en tanto que otros fueron “cerebrocéntricos”, la tesis que con el tiempo acabó desplazando a la anterior. Conviene recordar, con todo, que fue un convencido cardiocéntrico, William Harvey, quien en pleno siglo XVII descubrió y describió la circulación mayor de la sangre en un libro que ya desde el título, *De motu cordis*, recuerda su progenie cardiocéntrica.⁷

Pero el cerebrocentrismo fue imponiéndose poco a poco, como consecuencia de los datos aportados por la Medicina. El cerebro parecía ser la sede de la conciencia. La anatomía y la fisiología fueron aportando datos cada vez más convincentes sobre ello. La sede de la conciencia y de la propia vida, por tanto del alma, parecían asentar en el cerebro. Todavía Descartes, en pleno siglo XVII, especuló con que el punto de unión entre el cuerpo y la mente asentaba en la glándula pineal o epífisis.

Hoy no se duda de que el SNC es la sede de la conciencia, y que dentro de él hay una parte, la de aparición más reciente en el proceso evolutivo, el llamado neocortex o corteza cerebral, en que asientan las cualidades psíquicas específicamente humanas. Los neurofisiólogos suelen englobarlas en un término filosóficamente tan impreciso como “conciencia”, *consciousness*. De etapas previas de la evolución hemos heredado otras estructuras que constituyen el llamado paleocortex, que es condición necesaria pero no suficiente para la conciencia humana. Lo

⁷ Cf. D. Gracia, “Las razones del corazón”. *Naturaleza y Gracia* 2004;LI(2/3):345-392.

cual explica que en el ser humano haya una gran cantidad de actos, la mayoría, que no son conscientes en el sentido preciso del término. Son los actos inconscientes.

Del debate filosófico al debate científico sobre la inteligencia

Hasta hace relativamente poco tiempo, el debate sobre la inteligencia humana se hacía siempre por contraste con la llamada “inteligencia animal”, de tal modo que los protagonistas eran por lo general etólogos o psicólogos. Hoy el escenario es algo distinto, porque el contraste se establece con la llamada “inteligencia artificial”, debido a la importancia que en las últimas décadas ha cobrado la computación y esa nueva ciencia llamada informática.

Ahora la pregunta es si las máquinas son inteligentes, o si, volviendo al vocablo tan traído y llevado en el debate anglosajón, si los ordenadores serán capaces de tener “conciencia” o llevar a cabo procesos “conscientes”. La literatura sobre este tema se ha disparado en estos últimos años y parece que el debate no ha hecho más que comenzar.⁸ Ni que decir tiene que esto exige aclarar qué significa inteligencia o de qué estamos hablando cuando usamos esa palabra.

¿Qué es inteligencia? Primera respuesta: la inteligencia como capacidad de procesar información

Hubo que esperar al nacimiento de la Psicología experimental y a la aparición de los llamados tests de inteligencia, para que la idea de inteligencia cambiara de modo drástico. El criterio de inteligencia de la psicología no es filosófico sino operativo, práctico, y viene a identificarse con la “capacidad de procesar información”. Hay personas que tienen una enorme capacidad para procesar información numérica, o lingüística, etc. Esto llevó a distinguir varios factores de inteligencia, y luego uno general, el llamado factor g. Estos factores se miden mediante tests o pruebas en los que puede comprobarse la capacidad de procesar información de cada persona, que es distinta en los diversos factores y que converge en el llamado factor general de inteligencia.⁹

⁸ Cf. J. Monserrat, “El transhumanismo de Ray Kurzweil. ¿Es la ontología biológica reductible a computación?” *Pensamiento*, 71 (2015): 1417-1441.

⁹ Howard Gardner, *Teoría de las inteligencias múltiples*, 1983.

Con la aparición de los ordenadores, que también procesan información, y a veces en cuantía y velocidad muy superiores a las humanas, se ha comenzado a hablar también de “inteligencia artificial”. Los ordenadores, se dice ahora, son más o menos inteligentes, porque procesan información en medida mayor o menor.

¿Qué es inteligencia? Segunda respuesta: la inteligencia como capacidad de aprendizaje: la inteligencia animal

Hasta muy recientemente se ha venido pensando que los ordenadores tenían capacidad de procesar información, pero no de aprender, algo que sí hacen los seres vivos, por ejemplo, los animales. El aprendizaje de conductas es una cualidad que suele atribuirse al psiquismo, pero que es profundamente biológica. Los cuerpos vivos aprenden. Es lo que el fisiólogo Walter Cannon llamó *The Wisdom of the Body*.¹⁰ El organismo aprende a dormir en un cierto momento, a despertarse en otro, a tener hambre a una cierta hora, etc., etc. Los reflejos condicionados de Pavlov son un sistema de aprendizaje. En eso consiste la domesticación de perros, animales, etc. Y en cierto modo en eso consiste también la educación de las personas. Pero como veremos inmediatamente, quizá la educación no es mera domesticación. Alguno de los teóricos de la psicología del aprendizaje, como Skinner, creyó que con puros reflejos podría conseguirse el mundo feliz de Aldoux Huxley. Recuérdese su novela *Walden dos*.¹¹

Pero esto, que pasaba por ser específico de los seres vivos, el aprendizaje, resulta que ahora pueden hacerlo también las máquinas. Esto es lo que hoy se denomina “inteligencia artificial”. Su expresión más clara la encontramos en sistemas como el Chat GPT, hoy de rabiosa actualidad y que está causando tanta admiración como espanto.

¿Hay algo específico en la inteligencia humana?

La respuesta clásica

¹⁰ Walter Cannon, *The Wisdom of the Body*, New York, Norton, 1939.

¹¹ B F Skinner, *Walden Dos: Hacia una ciudad científicamente construida*, Barcelona, Martinez Roca, 1987.

Tras este rápido recorrido que hemos llevado a cabo, parece claro que por inteligencia se han entendido y se entienden cosas sumamente diversas. Desde el tiempo de los antiguos filósofos griegos, por inteligencia se ha venido entendiendo el *intus-legere o inter-legere*, la capacidad de abstraer las formas de la materia y captar la esencia. El *legere* remite a *lógos*, razón. Por eso al ser humano se le definía como *zôon lógon ékhon, animal rationale*, animal racional. El *intus legere* se entendía como la capacidad de “abstraer” la “forma” respecto de la “materia” de las cosas, y hacerse, de ese modo con las “ideas” de las cosas. Esta era la definición clásica, de progenia pletónica y aristotélica. En el interior del ser humano hay una realidad o sustancia de características completamente distintas a las materiales, que tiene la capacidad de trascender la materia y situarse en el universo de las ideas puras. En la naturaleza es posible que nunca haya habido un triángulo perfectamente equilátero, pero la mente humana puede pensar y razonar sobre el triángulo equilátero, con independencia de que tenga existencia empírica o mundanal. Por tanto, las ideas puras habitan en otro mundo distinto del material. Es el *hyperouránion tópos* de Platón, o lo que en toda la tradición occidental se ha entendido por la mente divina. De esto se deduce que la inteligencia es una capacidad o prerrogativa específicamente humana. Y más que humana, divina, de tal modo que en el ser humano no se da de modo completo, sino participado, como decía Platón. Dios es intelecto puro, *nóesis nóeseos nóesis*, dice Aristóteles,¹² y por ello Tomás de Aquino lo definirá como *actus purus*, dado que en él no puede haber nada en potencia.

La respuesta moderna

En la filosofía moderna, y de modo especial en el idealismo alemán, eso es lo que se denominará *Vernunft*, razón, a diferencia de la *Verstand*, entendimiento. Y de nuevo la *Vernunft* es condición divina, de la que el ser humano participa en alguna medida. En Dios no hay *Verstand* sino solo *Vernunft*.

Si acudimos a un representante paradigmático de la filosofía alemana del siglo XX, Heidegger, nos dirá que lo propio y característico del ser humano es la “comprensión” (*Verstehen, Verständnis*) del Ser, algo que no se da en los animales ni en las máquinas. Y con él toda la filosofía hermenéutica, al menos desde los tiempos de Dilthey.

¹² *Met XII 9: 1076 b 34.*

La respuesta de Zubiri

En el pensamiento español, Zubiri ha definido la inteligencia como la capacidad de actualizar las cosas en tanto que “realidades”. Y “cosa-real” es para él lo opuesto a “cosa-estimulo”, la que se agota en desencadenar una respuesta, que es el modo como parece que procede eso que se llama inteligencia animal, y por supuesto la inteligencia artificial. La inteligencia es una “formalidad”, la propia de la realidad humana, que permite “formalizar” o “actualizar” las cosas como “de suyo” o “en propio” reales, y por tanto como realidades y no como meros estímulos que suscitan respuestas. Si este es el sentido propio del término inteligencia, debería concluirse que, en el rigor de los términos, no es correcto hablar de inteligencia animal, y menos de inteligencia artificial.

Los datos científicos: Michael Tomasello

Los estudios de los investigadores de la Universidad de Leipzig, en especial Michael Tomasello, parecen indicar que lo específico de la inteligencia humana respecto a la de los otros primates es la capacidad de tener en cuenta en las decisiones a los otros, en vez de solo a uno mismo, como parece que sucede en los primates. De tal modo que el paso del modo “yo” al modo “nosotros” sería lo más específico y propio de la inteligencia humana.¹³

Pero sucede que esta enorme ampliación del concepto de inteligencia respecto de su sentido usual a lo largo de los siglos, se ha hecho, como no podía ser de otro modo, a costa de perder precisión. Como diría un lógico, ha ganado en extensión pero ha perdido en comprensión. Quien mucho abarca poco aprieta, según el refrán castellano. Si todo es inteligente, resulta que no vamos a poder definir con ello lo que es propio y específico

¹³ Cf. M. Tomasello, *Why We Cooperate*, Cambridge, Mass., MIT Press, 2009. *A Natural History of Human Thinking*, Cambridge, Mass., Harvard University Press, 2014. *A Natural History of Human Morality*, Cambridge, Mass., Harvard University Press, 2016. *Becoming Human: A Theory of Ontogeny*. Cambridge, Mass., Harvard University Press, 2019. *The Evolution of Agency: From Lizards to Humans*. Cambridge, Mass., MIT Press, 2022. Cf. también I. Pablo, Sahagún Kunhard y Gonzalo Alonso-Bastarreche, “Lo que nos hace humanos es la intencionalidad compartida: Michael Tomasello”, en: Juan Arana (Coord.), *Concepciones antropológicas de los protagonistas de la revolución neurocientífica*, Valencia, Tirant Humanidades, 2023, 361-389

de la realidad humana, la inteligencia humana, que es por donde comenzó esta historia.

Michael Tomasello piensa que es el “modo nosotros”. La tesis de Zubiri es distinta, aunque no incompatible: la aprehensión de las cosas como “realidades”. Habría que ver si ambos enfoques no están relacionados, y si el paso al modo “nosotros” no se debe, precisamente, a la capacidad de aprehender las cosas, no ya como estímulos que desencadenan respuestas, sino como cosas “de suyo” o “realidades”. De ser esto así, habría que concluir que lo específico de la inteligencia humana es, precisamente, esa capacidad de actualizar las cosas como cosas en sí, y no solo como estímulos que suscitan respuestas.

Otros datos científicos: Rodrigo Quian Quiroga

Este neurocientífico argentino es el descriptor de las llamadas “neuronas concepto”, expresión acuñada por similitud con la de “neuronas espejo” introducida por Rizzolatti unas décadas antes.¹⁴ Las neuronas concepto tienen su asiento en el hipocampo de los individuos de la especie humana pero no se encuentran en el de los animales, y para Quián Quiroga son las responsables de esa peculiaridad de nuestra especie que llamamos inteligencia específicamente humana, ya que el resultado de su actividad son conceptos, de modo similar a como las neuronas espejo reaccionan ante sentimientos. Las neuronas concepto no se caracterizan por responder a estímulos de los sentidos sino por codificar conceptos específicos. “Reflejan la percepción consciente y no el estímulo visual”.¹⁵ La tesis de Quian es que la inteligencia artificial podrá, quizá, igualar e incluso superar la inteligencia propia de los animales, pero que tendrá mucho más difícil imitar o remedar la actividad propia de las neuronas concepto. Este es el punto en el que estaría el salto cualitativo entre las inteligencias artificial y animal, y la inteligencia específicamente humana. De ser esto así, las neuronas concepto serían las responsables de lo propio y específico de la inteligencia humana.

De la cuestión formal a la pregunta por la génesis: el emergentismo

¹⁴ Giacomo Rizzolatti, Corrado Sinigaglia, *Las neuronas espejo: los mecanismos de la empatía emocional*, Barcelona, Paidós, 2006.

¹⁵ Rodrigo Quian Quiroga, *Cosas que nunca creeríais: De la ciencia ficción a la neurociencia*, Madrid, Debate, 2024.

Llegados a este punto, es lógico preguntarse por el origen de esta peculiaridad de la mente humana. De dónde procede esta peculiar *psykhé* del ser humano, ha recibido distintas interpretaciones a lo largo de la historia. Hay tres posibilidades interpretativas: la *animista*, que afirma la existencia de una sustancia espiritual, irreductible a la materia, en el ser humano, la *reduccionista* que reduce la mente a la materia orgánica, y la *emergentista*, que la interpreta como un salto cualitativo, y por tanto como una novedad estructural no reductible a sus componentes.¹⁶ Al menos desde el tiempo de Hegel se habla del “salto de la cantidad a la cualidad”.¹⁷ Por tanto, en la naturaleza hay saltos, y saltos que hacen aparecer cualidades nuevas, irreductibles a los elementos que las originaron. Esto hoy es tópico en química, y está en la base de teorías muy actuales, como la del caos.¹⁸ Pequeñas modificaciones elementales acaban produciendo saltos estructurales en que aparecen novedades inesperadas. Ya decía Aristóteles que el todo es más que la suma de sus partes. Y Zubiri habla, por ello, de unas “propiedades sistemáticas”, distintas e irreductibles a las “propiedades elementales”.¹⁹

¹⁶ Cf. Carlos Beorlegui, “El emergentismo humanista de Javier Monserrat”, *Pensamiento* 2015; 71 (269): 1165-1200.

¹⁷ Cg. G.W.F. Hegel, *Ciencia de la lógica*, Buenos Aires, Hachette/Solar, 1968, vol. 1, p. 323: “Todo nacimiento y muerte, en lugar de ser una gradualidad progresiva, son antes bien una interrupción de ella, y un salto desde la variación cuantitativa hacia la cualitativa”.

¹⁸ Edward N. Lorenz, *The Essence of Chaos*. Washington. University of Washington Press, 1995.

¹⁹ Cf. D. Gracia, “Zubiri en los retos actuales de la antropología”, en Pintor-Ramos (coord.), *Zubiri desde el siglo XXI*, Salamanca, Publ. Univ. Pontificia de Salamanca 2009, pp. 111-158. D. Gracia, “Problemas filosóficos de la génesis humana”, *Franciscanum*, 27:79 1985; 27(79): 17-32; Ibid. “Problemas filosóficos en genética y en embriología”, en F. Abel; C. Cañón (eds.), *La mediación de la filosofía en la construcción de la bioética*, Madrid, Federac. Internac. Universidades Católicas / Univ. Pont. Comillas 1993,, pp. 215-154; D. Gracia, “La antropología de Zubiri”, en J. A. Nicolás; O. Barroso (eds.), *Balance y perspectivas de la filosofía de Zubiri*, Granada, Comares 2004, pp. 87-116. J. Sáez, “Los retos actuales de la filosofía zubiriana y su ‘giro trascendental’ desde la noología a la trascendentalidad metafísica”, *Cuadernos salmantinos de filosofía* 2010; 37: 397-438.

Las tres hipótesis interpretativas citadas han dominado en diferentes épocas de la historia. La tradicional ha sido el “animismo”. El materialismo, especialmente en el siglo XIX, defendió muy beligerantemente el “reduccionismo”. Y en el siglo XX fue cobrando cada vez mayor fuerza la hipótesis emergentista.

La cuestión se reduce a saber si la inteligencia específicamente humana es una propiedad elemental o sistemática. La primera respuesta es la propia del creacionismo y la segunda la propia del emergentismo. Hoy por hoy no parece que haya razones que diriman completamente este dilema. Sí hay que decir que, aunque solo sea por permanecer fiel al principio de economía del pensamiento, parece más razonable la respuesta emergentista que la creacionista.

Según esta tesis, la emergentista, la inteligencia específicamente humana es una propiedad sistemática nueva, emergida de la complejidad estructural de los elementos constitutivos del sistema, que abre al ser humano a un mundo nuevo, el de lo que podemos llamar la vida del espíritu, o lo que Zubiri denominaría el orden de la realidad en cuanto realidad. Ese mundo nuevo es el mundo del espíritu.

Las hipótesis actuales

La neurofisiología lleva mucho tiempo investigando el modo como surge la conciencia específicamente humana. Hoy por hoy, no ha conseguido dar una explicación plausible que haya concitado la aprobación general y pueda darse por satisfactoria. Está más claro el modo o los modos como se pierde la conciencia, que el mecanismo por el que se gana. En lo que existe un acuerdo muy amplio, aunque desde luego no total, es en que ese milagro de la conciencia se consigue a través de la casi inextricable maraña de conexiones que se producen entre las neuronas del cerebro, y sobre todo de su corteza.

El año 1990, Francis Crick, el descriptor junto con James Watson de la estructura terciaria del ADN, publicó un libro titulado *The Astonishing Hypothesis: The Scientific Search For The Soul*, “La hipótesis sorprendente o revolucionaria: la búsqueda científica del alma”. Su hipótesis no es más que eso, una hipótesis, pero hoy aceptada por la mayoría de la comunidad científica. Dice que “las actividades mentales de una persona se deben por

entero al comportamiento de sus células nerviosas, células gliales y los átomos, iones y moléculas que las componen y las influyen.”²⁰ Lo de *astonishing* tiene que ver más con el marketing que con la realidad, y demuestra la ignorancia sorprendente por parte del laureado Crick del estado de la filosofía de su tiempo, ya que solo parece conocer, y poco, la de tiempos muy pasados, esa que somete a dura crítica. En cualquier caso, la idea hoy más aceptada es que el psiquismo específicamente humano es una propiedad sistemática dependiente de la estructura cerebral.

¿Podrá la inteligencia artificial alcanzar este grado de desarrollo?

El salto de la cantidad a la cualidad es un misterio que no sabemos muy bien cómo se produce. Lo que sí está claro es que su resultado es algo cualitativamente distinto a todo lo anterior. Eso es lo que llamamos un salto cualitativo. Se produce en las combinaciones químicas, en las que de elementos con determinadas propiedades surgen compuestos con propiedades nuevas, irreductibles a las anteriores. En la evolución humana, parece que ese salto también se ha producido, y probablemente no una vez, sino muchas. Zubiri se ocupó de esto, hace ahora más de medio siglo, en su artículo “El origen del hombre”.²¹ La inteligencia humana es cualitativamente distinta de la animal e irreductible a ella, pero a la vez dependiente de ella. Surge, como decía Zubiri, por una especie de “desgajamiento exigitivo”, que a fin de cuentas es un tipo de emergencia.

¿Podría suceder algo similar en los procesadores de información que utiliza la inteligencia artificial? ¿Podría darse en ellos el salto de la cantidad a la cualidad, y por tanto el salto a la inteligencia específicamente humana? No parece que haya argumentos para negar esa posibilidad, que ya se ha dado en el mundo animal. ¿Podrían llegar a los ordenadores a ser inteligentes, en el sentido preciso y específico o propio de la inteligencia humana? Podría ser. No hay razones para negar esa posibilidad.

²⁰ Francis Crick, *La búsqueda científica del alma: Una revolucionaria hipótesis para el siglo XXI*, Barcelona, Debate, 1994, p. 345.

²¹ Cf. X. Zubiri, “El origen del hombre”, en *Escritos menores (1953-1983)*, Madrid, Alianza, 2006, 65-102.

Pero eso aún no ha sucedido y no sabemos si sucederá. Lo que sí podemos decir es que ese salto a día de hoy no se ha producido, y que esto que se llama en la actualidad inteligencia artificial, no lo es en el rigor de los términos, es decir, cuando la palabra inteligencia se entiende en sentido fuerte, y no como mera capacidad de procesar información o simple capacidad de aprendizaje.

Conclusión

De ser un puro asunto de especialistas, el tema de la inteligencia artificial ha pasado a ocupar y preocupar a la opinión pública, hasta convertirse en uno de los temas y problemas más sorprendentes, a la vez que preocupantes, de nuestra actualidad. La sorpresa del fenómeno ha dado lugar a todo tipo de especulaciones sobre lo que puede llegar a suceder en el próximo futuro. De hecho, ha puesto en guardia a las principales instituciones internacionales, que han comenzado a plantearse cómo controlar este nuevo y revolucionario fenómeno, preocupadas por no llegar tarde a su necesaria regulación y control. De lo que no parece haber duda razonable es de que promete ser una de las mayores revoluciones técnicas acaecidas en la historia entera de la humanidad, llamada a transformar muy profundamente la vida de nuestras sociedades. De ser esto así, la reflexión sobre ella es muy razonable pensar que habrá de ocuparnos de modo cada vez más intensivo a lo largo de los próximos años, quizá décadas.

En estos inicios del proceso, la primera pregunta que cabe hacerse es la referida a lo que los escolásticos llamaban el *quid nominis*. ¿Es correcta la expresión inteligencia artificial? ¿Qué significa exactamente? ¿Sabemos con precisión de qué estamos hablando?

La palabra inteligencia procede de la raíz latina *lego* que significa leer, por lo que originariamente tiene el sentido de leer dentro o captar la esencia o realidad de algo. En la filosofía antigua, esto acontecía mediante el proceso mental de abstracción de la llamada forma respecto de la materia, de tal modo que entender algo era captar con la mente lo que ella es, eso que los filósofos griegos llamaron su forma, abstrayéndola de la materia. Al decir que algo es una silla, la materia de la silla no penetra en nuestra mente, pero sí su forma, eso que nos hace decir que es una silla y no cualquier otro objeto. Esa captación de la forma, que viene a

identificarse con la esencia de las cosas, en este caso de la silla, es la que expresamos atribuyéndola un nombre, silla. El lenguaje es, según esto, la expresión de esa esencial formal de las cosas. Los griegos utilizaron una palabra, *lógos* para designar este sorprendente fenómeno, y de ahí que ese término signifique en griego tanto razón o intelección como palabra. Y Aristóteles definió al ser humano, a diferencia de los demás seres vivos, como el animal dotado de *lógos*, es decir, de razón y de palabra, lo que le distingue de todos los demás animales.

Esta es la definición de inteligencia que ha circulado a lo largo de toda la historia de la cultura occidental, hasta hace relativamente poco tiempo, en que la ciencia que se ocupa del estudio del comportamiento animal, la etología, empezó a hablar de “inteligencia animal”, algo incomprensible en toda la tradición anterior. Mucho más reciente es aún la expresión “inteligencia artificial”, nacida con el auge de la informática en el último medio siglo. Y surge la pregunta: ¿cabe hablar de inteligencia a propósito de las máquinas? ¿Son las máquinas inteligentes?

De acuerdo con la definición griega de inteligencia, evidentemente, no. Las máquinas, al menos por ahora, no son capaces de llevar a cabo el proceso de abstracción de que hablaron los filósofos antiguos, ni tampoco de aprehender las cosas como realidades. Su gran habilidad está en el proceso y manejo de la información, de tal manera que las máquinas son inteligentes si por inteligencia entendemos, como hoy es usual, la capacidad de procesar información. En este sentido amplio, amplísimo del término, no hay duda de que los ordenadores son inteligentes, y los animales también. Es más, poco a poco vamos asumiendo el hecho de que, así definida la inteligencia, los animales, y sobre todo las máquinas, son en ciertos tipos de funciones más inteligentes que los propios seres humanos.

Pero inmediatamente surge la objeción de que las máquinas no tienen conciencia de lo que hacen, ni por tanto pueden salir responsables de sus actos. Tanto de los animales como sobre todo de ellas nos separa, entre otras características, la ética. No parece que las máquinas tengan conciencia del deber. ¿Podrán adquirirla en algún momento? No lo sabemos, pero tampoco podemos descartarlo completamente. En la naturaleza son continuos los saltos de la cantidad a la cualidad, de tal manera que de la complejización de estructuras materiales acaban

surgiendo propiedades no solo nuevas sino cualitativamente distintas a las de las anteriores e irreductibles a ellas. Es la teoría que se conoce con el nombre de “emergentismo”. Sin ella es imposible entender la evolución de los seres vivos, en la que los saltos de la cantidad a la cualidad parecen haber sido muchos. ¿Llegarán algún día las máquinas a ser inteligentes, no en el sentido amplio y actual de ese término, el de capacidad de procesar información, sino en el sentido de tener conciencia de lo que hacen? ¿Podrá darse ese salto? No lo sabemos. Lo que sí sabemos es que por ahora ese salto no se ha dado y que incluso en esta época de euforia y optimismo ante los resultados de la inteligencia artificial, el momento parece lejano. Pero tampoco puede descartarse.

De la teoría a la práctica: reflexiones desde la ética

La inteligencia humana tiene como nota fundamental el ser “proyectiva”. El proyecto es siempre de futuro, razón por la cual Ortega y Marías decían que el ser humano es “futurizo”, está inexorablemente lanzado hacia el futuro. Vive desde el futuro. En esto consiste lo que Ortega llamaba su condición “proléptica”.

El término del proyecto es la puesta en práctica de lo proyectado, y por tanto su realización. Entre todas las opciones posibles, se elige una que, por las razones que sean, se considera óptima. De tal modo que la decisión humana consiste siempre en la elección de una entre las posibilidades diversas que se le presentan al individuo en cada momento.

La naturaleza ofrece “recursos” al ser humano, que pocas veces son ya en sí mismos “posibilidades” de vida. Por lo general, los recursos naturales necesitan ser transformados a través de la inteligencia humana para convertirse en posibilidades reales de vida. Un sencillo ejemplo aclarará esto. Todo el arte culinario consiste en hacer aptos para la alimentación humana productos que en principio no lo son. Esos productos más o menos naturales son meros recursos, que solo a través de la técnica culinaria se convierten en verdaderas posibilidades de vida.

La conversión de los recursos naturales en posibilidades humanas se hace a través de la ciencia y la experiencia humanas, por tanto, de la ciencia y de la técnica. Cuanto más avanzan éstas, mayor es nuestra capacidad de transformar recursos en posibilidades de vida, de tal modo que el abanico

de estas últimas se amplía progresivamente, hasta cobrar una amplitud insospechada.

Es obvio que, cuanto mayor es esa amplitud, más posibilidades de elección tenemos los seres humanos, y por tanto más compleja se vuelve la toma de decisiones. Caso de que no hubiera más que una posibilidad, ésta vendría a identificarse con el recurso, y el proceso de elección se convertiría en lineal, y por tanto automático. Pero cuanto mayor es el abanico de posibilidades, más compleja resulta, en principio, la elección.

Digo en principio, porque en la práctica hay elementos que pueden simplificar grandemente el proceso. Estos son muchos y de muy distinto tipo. Así, la adicción, la costumbre, el hábito, la prisa, etc., hacen que tomemos decisiones en cortocircuito, de modo automático o casi. Todos los citados son procedimientos de elección que se caracterizan por priorizar factores que no están directamente relacionados con la optimización objetiva de los resultados. Si se elige por gusto, por comodidad, por prisa, por hábito compulsivo, es evidente que se está haciendo por motivos que no son el que en principio consideramos adecuado, el elegir la opción que por sus características propias pueda considerarse la mejor o la óptima. Es decir, nuestra decisión sería subóptima, por estar condicionada por ciertos “sesgos”.²²

Las decisiones son tanto más racionales o humanas cuanto menos están determinadas o condicionadas por “sesgos”, los llamados sesgos de decisión. Estos son muchos, y además últimamente son objeto de gran atención por los psicólogos, ya que su manejo inteligente por parte de personas o grupos permite manipular la opinión, tanto pública como privada, y de ese modo conseguir que las personas hagan lo que nosotros queremos que hagan, pero creyendo que hacen lo que ellos quieren. Esto es lo que el sociólogo y filósofo Max Weber definió como *Herrschaft*, “dominación”. En esto se basa buena parte de la teoría del *marketing* y de lo que hoy se conoce con el nombre de *behavioral economics*. Los libros de Kahneman y Thaler son buena prueba de ello.²³ Contra la dominación no

²² Cf. Diego Gracia, “Problemas con la deliberación”, *Folia Humanistica* 3; 2016: 1-16.

²³ Daniel Kahneman, *Pensar rápido, pensar despacio*, Barcelona, Debolsillo, 2013. Cass R. Sunstein & Richard Thaler, *Un pequeño empujón*, Madrid, Taurus, 2009.

hay más que un antídoto, tan viejo como la propia filosofía. Se llama “deliberación”. Una de nuestras más graves asignaturas pendientes.²⁴

La deliberación no es en principio un asunto de economía o de sociología sino de lógica y de ética. De lo que se trata es de saber no cómo se utiliza o “puede” utilizarse de hecho la IA, sino cómo “debe” hacerse. Para evitar sus riesgos, una posibilidad sería la prohibición absoluta. Ni que decir tiene que tal alternativa es irreal además de imposible. Otra posibilidad es la de su regulación. Es lo que se está intentando por parte de algunos organismos, como la Unión Europea. Parece que están topando con un problema, y es que aún no se conocen bien las posibilidades de la IA, razón por la que resulta muy difícil prohibir o regular aquello que aún no se conoce o es poco y mal conocido.

La ética no es muy partidaria de las prohibiciones. Fía más de la educación en la responsabilidad. De lo que se trata es de saber cómo fomentar el uso responsable de estas tecnologías, de tanto potencial positivo como destructivo y adictivo. Esta tiene que ser la gran propuesta ética, la educación en la responsabilidad.

No es la primera vez que nos enfrentamos con retos de este tipo. Pensemos en el consumo de drogas y sustancias adictivas. Su prohibición no ha impedido la difusión masiva de su tráfico y consumo. No parece, pues, que la mera prohibición legal sea suficiente. ¿Cómo educar a nuestros hijos en este tema? ¿Nos contentamos con que los gobiernos prohíban su tráfico, o nos esforzaremos porque los ciudadanos sepan lo que tienen entre manos e busquemos cómo promover y fortalecer la responsabilidad de las personas, sobre todo de los jóvenes, a fin de que sean prudentes en las decisiones que toman?

Cuanto mayor es el abanico de posibilidades de decidir que tenemos los seres humanos, sobre todo aquellos más jóvenes y con menor experiencia de la vida, más necesario resulta educarlos en la gestión seria y responsable de sus decisiones. Un joven de dieciséis años tiene hoy más posibilidades a su alcance, tanto positivas como negativas, que tuvieron nunca a lo largo de toda su vida sus mayores. De ahí la perentoria necesidad de educarlos en el autocontrol y la gestión responsable de sus

²⁴ Cf. Diego Gracia, *El animal deliberante: Teoría y práctica de la deliberación moral*, Madrid, Triacastella, 2024.

decisiones. Dicho de otro modo, hoy se requiere que los jóvenes sepan gestionar de un modo más autónomo que nunca antes sus decisiones y su vida.

Estos comentarios pueden parecer en exceso abstractos. Pero no resulta difícil concretarlos. Piénsese, por ejemplo, en los estadios del desarrollo del razonamiento moral de Kohlberg. Según ellos, los adolescentes están en la fase de desarrollo convencional, de la que muy pocos, según los estudios empíricos, van a pasar a lo largo de su vida al estadio posconvencional. Esto es lo que cada vez se hace necesario y urgente remediar. Cuanto más amplio es el abanico de posibilidades a su alcance, más madura y responsable necesita ser su capacidad de elección. Solo las personas verdaderamente posconvencionales están adecuadamente preparadas para utilizar de modo responsable tecnologías como la IA. No es el único caso. En pareja situación se encuentra el acceso a las drogas y demás productos adictivos, consumo de pornografía, etc.

Potter introdujo en 1970 el término “bioética” para significar que el progreso de la tecnociencia debía ir parejo al de la reflexión moral y viceversa, a fin de evitar desastres que pueden llegar a ser apocalípticos. La reflexión ética debe acompañar el progreso científico e ir a su mismo ritmo. No se trata tanto de prohibir cuanto de gestionar nuestros actos de modo crítico y responsable, es decir, posconvencional. Algo que dista mucho de haberse conseguido.

Es difícil no recordar en esta situación a Hans Jonas y su “heurística del miedo”.²⁵ Él acuñó esa expresión a propósito de las armas nucleares y de la crisis ecológica. Pero vale también para el tema que ahora nos ocupa, el de la IA. Tiene un gran potencial positivo, pero también negativo. Como ya dijera Kant, “los preceptos que sigue el médico para curar perfectamente al hombre y los que sigue el envenenador para matarlo, seguramente son de igual valor, en cuanto que cada uno de ellos sirve para realizar cumplidamente su propósito.”²⁶ Y Zubiri recordaba entre nosotros que el progreso científico-técnico no solo permite incrementar los bienes sino

²⁵Hans Jonas, *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica*, Barcelona, Círculo de lectores, 1994.

²⁶ Immanuel Kant, *Fundamentación de la metafísica de las costumbres*, Akad. IV, 415, 16-19.

también, y en proporción similar, los males.²⁷ Todo depende de que se use o no responsablemente. El problema es que resulta más sencillo y cómodo utilizar mal estas técnicas que hacerlo bien.²⁸ Esto es lo que se ha dado en llamar el “X-Risk” o *existential risk*.²⁹ Para gestionarlo adecuadamente no hay más que un medio, la educación en la responsabilidad; es decir, la ética.

²⁷ Xavier Zubiri, *Sobre el sentimiento y la volición*, Madrid, Alianza, 1992, p. 284.

²⁸ Cf. Michal Kosinski, Poruz Khambatta, Yilun Wang, “Facial Recognition Technology and Human Rates Can Predict Political Orientation From Images of Expressionless Faces Even When Controlling for Demographics and Self-Presentation”, *American Psychologist*. Advance online publication. <https://dx.doi.org/10.1037/amp0001295>.

²⁹ Caesar Alimsinya Atuire, Jean-Christophe Bélisle-Pipon, Vardit Ravitsky, Anina Ho, Nancy S Jecker, “AI and the falling sky: interrogating X-Risk”, *J Med Ethics*. 2024 Apr 4:jme-2023-109702. doi: 10.1136/jme-2023-109702. Online ahead of print.