

LA INTELIGENCIA ARTIFICIAL Y
LA REALIDAD RESTRINGIDA:
Las estrecheces metafísicas de la tecnología

José Luis González Quirós
Universidad Rey Juan Carlos, Madrid

Resumen: La visión de la realidad queda restringida cuando pensamos únicamente en un universo de esquemas, de máquinas, cuando identificamos nuestra conciencia con alguna clase de objetos, una ficción que crea una imagen innecesariamente restrictiva de lo que hay, un reduccionismo de la realidad que no se deduce de la tecnología misma, sino de la mala metafísica que se quiere hacer en su nombre.

Palabras clave: IA, máquinas, reduccionismo, metafísica, realidad, existencia, conciencia, inteligencia.

Artificial Intelligence and Restricted Reality: The Metaphysical Constraints of Technology

Abstract: The vision of reality is restricted when we think only of a universe of schemes, of machines, when we identify our consciousness with some kind of objects, a fiction that creates an unnecessarily restrictive image of what there is, a reductionism of reality that does not come from technology itself, but from the bad metaphysics that is done in its name.

Keywords: AI, Machines, Reductionism, Metaphysics, Existence, Consciousness, Intelligence.

Recibido: 27/02/2019 **Aprobado:** 20/02/2019

Cualquier forma de caracterizar la época que estamos viviendo ha de fijarse en el extraordinario desarrollo de tecnologías radicalmente contemporáneas, es decir sin apenas antecedentes. Creo que por mucho que se examine el pasado no será fácil encontrar nada comparable a esta *era digital*. Una cualidad que llama poderosamente la atención en relación con sus novedades es que apenas haya discrepancia acerca de sus beneficios, digamos, materiales, pero abundan, en cambio, las dudas sobre su impacto intelectual y moral. Es un lugar común reconocer que el enorme incremento de la cantidad de información disponible está creando formas nuevas de

ignorancia y de manipulación, que se está derribando casi cualquier frontera concebible entre la realidad y la ficción. Uno de los campos en los que esa *con-fusión* está adquiriendo enorme notoriedad es el de la acelerada convergencia de lo artificial y lo real, de lo informático y lo vivo, lo que, en la jerga al uso, se conoce con el nombre de *ciberrealidad*, una peculiar mezcla de posibilidades, invenciones y fantasías en cuyo entorno no dejan de producirse sorpresas, aunque a veces pertenezcan más al reino de la propaganda que al de las verdades sin mácula.

Unas tecnologías tan poderosas como las que se están desarrollando ante nuestros ojos suscitan casi inevitablemente cuestiones de carácter metafísico, pues ponen en juego la condición misma de lo real. No en vano transforman la naturaleza en un universo plenamente tecnológico y lo esconden tras varias capas *virtuales*, dando lugar a un mundo que muchos consideran extrañamente humano, en el que nada parece imposible, pues los límites se rebasan de manera continua, y eso invita a no reprimir los deseos, a sentirse como dioses, a considerar irrelevante todo lo que no se pueda entender, o a manejarlo de cualquier manera cubriéndolo de palabrería. En ese tipo de escenarios se suele hablar, con la misma confianza con que se afirma que el cielo es azul, de la inexistencia de la conciencia, de la identidad entre mentes y máquinas y de la completa conectividad entre cerebros y sistemas, la realización efectiva de una nueva *síntesis* que Ray Kurzweil ha bautizado como *singularidad*. El objetivo de estas líneas es analizar algunas de las implicaciones filosóficas de estos proyectos partiendo del desarrollo de la disciplina en cuyo ambiente han surgido, de la Inteligencia Artificial (en adelante IA).

En la filosofía contemporánea, al menos en esa parte de ella que exhibe sus rigores con el lenguaje, parece existir una amplia tolerancia hacia la tontería, lo que puede que suceda en homenaje a una virtud intelectual, la apertura a nuevas ideas, pero que, al tiempo, hay que considerar también como una consecuencia del hecho incontestable de que, en el mundo del que la filosofía es solo una parte, y al que ha de estar abierto, las tonterías gozan ahora de un prestigio y un respeto sin igual en anteriores épocas. Esta

reflexión me la ha sugerido un texto de Galen Strawson (2018) donde afirma que negar la existencia de la conciencia es la más grande de las tonterías nunca defendidas. El joven Strawson se pone en la estela de Cicerón y asevera que los filósofos somos extremadamente propensos a defender tontunas. Al leerlo he pensado que se trata de un defecto difícil de encontrar en los clásicos, supongo que no tanto en Roma donde los filósofos nunca fueron multitud ni de primera clase, pero, efectivamente, resulta sospechosamente frecuente entre los contemporáneos.

Curiosamente, querría subrayarlo, han sido los llamados filósofos analíticos los que, en esta clase de temas, han sido más proclives a decir alguna enormidad, y a esconderlas con cierta habilidad en textos impenetrables, lo que llama especialmente la atención, pues cuando se ponen a hacer distinguos se quiebran de puro sutiles. Es llamativo que esa preocupación por lo que se puede decir y lo que no puede decirse, no haya evitado sostener algunas liviandades e inconsecuencias, por ejemplo, defender la idea de que la mente pueda ser algo como *software* o una *máquina virtual*, olvidando que el cerebro, o más en general, el cuerpo que la *incorpora*, es una realidad biológica que experimenta de manera continuada diversas alteraciones¹, algo vedado a cualquier sistema que procese un programa, pero también resulta significativo que tantas cautelas *lógicas* no hayan sido demasiado eficaces para extender a otros ámbitos una cierta prudencia respecto a lo que se dice.

Los filósofos no cesamos de darle vueltas a ese tipo de cuestiones, entre otras cosas por tener, al menos en España, el hábito de pensar que si lo ha dicho Fulano o Zutano, que es un sabio extranjero, *deben ser ciertos los toros*. Verdad es que, pese a que lleve cerca de cincuenta años dedicado a leer muchas de las cosas que se escriben respecto a lo que se suele conocer como *Filosofía de la Mente*, estoy dispuesto a asegurar que hace mucho, muchísimo, que no leo ni puedo pensar algo nuevo al respecto, lo que tal vez explique la previsibilidad, monotonía y pesadez de muchos filósofos. Algunos, sin

¹ Pueden verse a este respecto las críticas de Gelernter (1994: 123) a Ned Block y al propio Dennett.

embargo, consiguen hacerse un nombre con técnicas tan escasamente persuasivas. Tal vez el truco consista en decir cosas que puedan parecer extravagantes o *revolucionarias* a los que no son del gremio, porque la revolución es lo único que, al parecer, soporta las tradiciones por los siglos de los siglos. Sartre se anunciaba hace muchas décadas diciendo que preparaba un libro de *filosofía destructiva*, y ahora los hay que, pese a llevar un supuesto rigor lógico por bandera, consiguen colocar las mismas ideas una y otra vez, aunque es justo reconocer que les han salido unos competidores muy duros entre los gurús de Silicon Valley, con el agravante de que, además de defender extraños filosofemas como cualquier académico de moda, se hacen millonarios por otros procedimientos.

Todo lo cual se dice para pedir excusas por no ser capaz de sugerir ideas muy distintas a las que he defendido en ocasiones similares, y mira que me disgusta repetir la lección. Pero si en algo estoy de acuerdo con Quine es en reconocer que las novedades en filosofía están extraordinariamente caras, así es que me disculparán, pero no espero ni aspiro a sorprenderles. Aparte de las dificultades intrínsecas del caso y el hecho de que llevemos unos miles de años discutiendo sobre asuntos que a veces puede que no entendamos del todo, una explicación adicional podría residir en que, según se ha dicho desde casi siempre, el origen de la filosofía está en el asombro, pero vivimos en una época escasamente capaz de asombrarse, que, en realidad, finge el asombro ante supuestas novedades, tecnológicas o no, que casi siempre son lo mismo, pero ha perdido la capacidad de asombrarse de cosas bastante más básicas, como que el sol salga y la lluvia caiga. La culpa de esto, me parece, la tiene, en cierto modo, la ciencia moderna o, más bien, mucho de lo que se nos ha dicho a su propósito, es decir ciertas filosofías que se presentan como la interpretación correcta de la ciencia y que suponen una especie de *super-saber* para el cual ya no queda nada de lo que asombrarse porque todo está, en lo sustancial, explicado. Como contraste, recuerden aquello de Fray Luis de León en la *Oda a Felipe Ruiz*,

Veré las inmortales / columnas, do la tierra está fundada, / las lindes y señales
con que a la mar hinchada / la providencia tiene aprisionada. / Por qué tiembla

la tierra; / por qué las hondas mares se embravecen; / dó sale a mover guerra / el Cierzo; y por qué crecen / las aguas del Océano, y descrecen. / De dó manan las fuentes; / quién ceba y quién bastece de los ríos / las perpetuas corrientes; / de los helados fríos / veré las causas, y de los estíos. / Las soberanas aguas / del aire en la región quien las sostiene; / de los rayos las fraguas / do los tesoros tiene / de nieve Dios; y el trueno dónde viene.

Estas palabras están escritas antes del XVII, antes que Galileo y Newton, y su ingenuidad puede provocar sonrisas, pero tal vez deberían hacer que nos preguntemos cómo el *desencantamiento del mundo* que ha traído la ciencia no se ha trasladado a la tecnología, puesto que la tecnología, según dijo Clark, cuando es avanzada es indistinguible de la magia, y, si se trata de magia, es evidente que se puede andar sobre las aguas o separar un cuerpo en dos y andar con la cintura, todos lo hemos visto en YouTube alguna vez. El desarrollo conjunto de la ciencia y de la tecnología nos han acostumbrado a lo inverosímil y a lo *imposible*, y puede ser que esa frontera entre lo que es posible y lo que no lo es la situemos mal una y otra vez, pero es excesivamente aventurado dar por hecha su inexistencia.

Vayamos a nuestro tema, pero queda advertido que la discusión que se hace es inseparable de ciertos paradigmas culturales, por llamarlos de algún modo, de modo que si yo declaro paladinamente que soy una especie, presuntamente civilizada, de *dualista*, tengo que reconocer el *derecho* a la existencia de los reduccionistas, los funcionalistas, los emergentistas, los pansiquistas, como lo es, en cierto modo el propio Galen Strawson, por la sencilla razón de que ninguno de nosotros podemos salir de este mundo y verlo como realmente es, sin nuestra preciosa y privilegiada presencia, según esa magnífica caricatura que hizo Woody Allen en *Annie Hall* sacándose de la manga a Marshall McLuhan en persona para meterle en el plano de esa película con el fin, escasamente caritativo, de desmentir al presuntuoso profesor de la Universidad de Nueva York que estaba diciendo bobadas sobre las ideas de McLuhan y el cine de Fellini.

1. La inteligencia artificial, una promesa permanente

El nombre y el origen mismo de la IA como nueva disciplina, se debe a John McCarthy que usó ese término por primera vez en 1956 en un seminario en Dartmouth College en el que también participaron Allen Newell, Herbert Simon, Marvin Minsky y Arthur Samuel. La IA pudo ponerse en marcha, especialmente gracias a una doble serie de avances en las décadas anteriores, y, en el fondo, al impulso tecnológico de la guerra. Por una parte, el desarrollo, en Inglaterra y en los EE. UU., de los primeros sistemas de cálculo basados en tubos de vacío, y, por otra, el trabajo teórico de Turing que, ya en 1935, había descrito lo esencial de una máquina abstracta capaz de almacenar memoria y de operar símbolos conforme a diferentes tipos de instrucciones virtuales. Sin embargo, el desarrollo definitivo vino con el trabajo de Von Neumann y otros en el laboratorio de Los Álamos a comienzos de los años cincuenta del pasado siglo. Von Neumann, que conocía los trabajos de Turing y le había ofrecido trabajar con él como *postdoc*, y ha reconocido la importancia del inglés en la creación de los conceptos básicos de teoría de la computación, ha sido acusado en ocasiones de apropiarse de ideas de Turing, pero le dio a la idea básica una formulación mucho más elaborada de forma que, en general, se le reconoce que fue quien sentó las bases definitivas para el desarrollo de la industria informática que está a la base de todos los desarrollos de la IA.

Tras los primeros éxitos, recibidos con gran expectación en una atmósfera de optimismo tras los años de la guerra, los fundadores de la disciplina fueron tal vez excesivamente ingenuos, sobreestimaron las capacidades de cálculo disponibles y, todavía más, las dificultades de todo tipo que surgirían al tratar de implementar funciones cognitivas que se nombran con términos fáciles de utilizar, como *ver*, *percibir*, *comprender*, *razonar*. En los primeros años, Minsky pensaba, por ejemplo, que hacer una máquina capaz de mantener una conversación de amigos en el trabajo sería posible en breve, y H. Simon afirmó que en veinte años las máquinas podrían hacer

cualquier cosa que pudiéramos hacer las personas. No ha sido así, es evidente. Pronto se vio que el problema era algo más complicado, y se abrió paso la idea de que como escribió David Gelernter (1996:229), “La ciencia informática es un campo lleno de chiflados ávidos de novedades”.

Por tanto, puede decirse que las promesas *filosóficas* de los fundadores, la creación de máquinas tan capaces y flexibles como los seres humanos, todavía no se ha conseguido, y cabe dudar de que ese camino lleve a parte alguna, aunque siguen existiendo informáticos que, además de ser excepcionales en lo suyo, continúan suponiendo que ese panorama se cumplirá a no mucho tardar. Ahora mismo el representante más característico de esa tendencia es Raymond Kurzweil, autor, por ejemplo, del programa que permite a los ordenadores leer textos en *pdf* en voz alta, y director de ingeniería de Google, quien sostiene que la comunicación directa entre ordenadores y cerebros humanos, lo que él ha llamado la *singularidad*, será una realidad a muy corto plazo.

Un caso relevante es el de los debates y confusiones en torno a las capacidades cognitivas de los sistemas de IA, que, en buena parte, pueden considerarse como una serie de equívocos que tal vez deriven de haber llamado *Inteligencia* a algo que bien podría denominarse como *Automática*, pues resulta verosímil que algo más de modestia al bautizar los inventos nos hubiese ahorrado algunas querellas sobre el particular. Por ejemplo, Judea Pearl, una de las figuras relevantes en el desarrollo de la IA, afirma que mientras las máquinas no aprendan a manejarse con el esquema causa-efecto, lo que equivale a una forma esencial de entendimiento, mientras funcionen como meras aproximaciones a curvas de probabilidad, su inteligencia será extremadamente insuficiente².

Pero, de la misma manera que las afirmaciones gruesas sobre *máquinas que piensan* han sido genéricamente abandonadas, se abre paso la idea de que algunos de los sistemas están superando claramente el nivel de meros calculistas porque, según sus creadores, la IA está empezando a conseguir

² Puede verse una entrevista con Pearl en <https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/>.

resultados que, además de exceder la capacidad humana de cálculo, cosa que ya era muy sabida, parecen haber superado los límites impuestos por una simple rutina de seguimiento de reglas.

Éste es el panorama que se puede vislumbrar tras los éxitos de *AlphaZero*, el algoritmo de *aprendizaje profundo* de *Deep Mind*, la empresa de IA de *Alphabet*, que, ha derrotado a todos los programas existentes de ajedrez (y de otros juegos asimilables) y no lo ha hecho con la característica *brutalidad* de los programas clásicos, sino con formas de jugar que antes parecían reservadas a los seres humanos y que, según sus creadores, proporcionan *descubrimientos*, maneras nuevas de entender y jugar al ajedrez muy alejadas de las rutinas propuestas por los programadores. Si es verdad que *AlphaZero* ha sido capaz de comprender y utilizar creativamente la *esencia* del ajedrez (que según Bobby Fisher consiste en *la maldad*), de emplear estrategias que, hasta ahora, no se conocían, jugando no como lo hacían las *máquinas* sino como lo haría un consumado maestro, es evidente que estaremos ante una novedad muy estimulante en la IA, ante el desarrollo de formas superiores de inteligencia y aprendizaje. Sobre este punto pueden verse las interesantes reflexiones de Steven Strogatz, un matemático de Cornell, que sugiere que esta nueva generación de algoritmos puede estar llevándonos a descubrir un nuevo tipo de *visión* o de *inteligencia* de carácter casi *oracular*, a demostrar verdades cuya plena validez solo podamos contrastar a posteriori, como, según Strogatz, ha sucedido ya con la demostración del *teorema de los cuatro colores*, pero de tal manera que la comprensión de tal hallazgo no esté al alcance de una inteligencia humana. Estaríamos ante formas de inteligencia, y no de mero cálculo, aunque más allá de la capacidad de nuestra conciencia³, de nuestra inteligencia consciente, lo que implicaría que tendríamos que enfrentarnos con dificultades nada pequeñas para entender cuál ha sido precisamente el *método* seguido por las máquinas para obtener sus resultados.

³ Un resumen de su libro *How Calculus Reveals the Secrets of the Universe*, cuya aparición se anunció para abril de 2019, puede verse en <https://www.nytimes.com/2018/12/26/science/chess-artificial-intelligence.html>.

Sea lo que fuere de ello, es evidente que los avances en IA no restringen ninguna realidad, son las cosas que algunos comentaristas dicen al respecto las que insinúan una realidad demediada, restringida, absurdamente desposeída de algo que cualquier persona sin demasiadas entendederas ni lecturas vería como algo absolutamente obvio: que existe la conciencia, que es una propiedad muy singular y peculiar de la vida y que, obviamente, no parece razonable suponer que las máquinas tengan conciencia, por muy sofisticadas que sean, y por muchos éxitos que supuestamente alcancen al superar la *prueba de Turing*. Confundir inteligencia y conciencia no es muy recomendable: toda conciencia ha de tener alguna forma de inteligencia, pero no toda forma de intelección ha de ser forzosamente, y en todo momento, consciente, cosa que admitiremos con facilidad a nada que reparemos en que muchas veces encontramos la solución verdadera a un problema sin saber cómo hemos llegado exactamente a encontrarla. Nuestro cerebro ha actuado inteligentemente, pero no hemos sido conscientes de lo que ha hecho.

La era digital ha desarrollado con éxito una tecnología en la que se potencia conjuntamente el éxito de dos estrategias, la vocación atomista del reduccionismo y la creación de máquinas virtuales, pero no es razonable olvidar su distinto fundamento y su respectivo valor: reducir a lo más simple permite determinar los elementos que componen una realidad dada cuya articulación pueda explicar determinados fenómenos, mientras que la digitalización se construye no para mejor conocer la realidad, sino para manejar con eficacia formas de significado previamente dadas, que no se descifran, sino que se cifran digitalmente. La electrónica y la digitalización han creado maravillas conjuntas, pero su éxito se basa en consideraciones distintas.

Lo que en la actualidad hacen la mayoría de los ordenadores, es algo que no hacía ninguno en los años cincuenta, interactuar de manera habitual con usuarios que pueden ignorar todo sobre su funcionamiento. Que podamos hacer con ellos infinidad de cosas, tales como *hablar, escribir, ver*, etc. constituyen *aplicaciones* enteramente imprevistas por los fundadores del campo, como se ve muy bien cuando se cae en la cuenta de que el propio John Von

Neumann, que en 1955 pensaba que en unas décadas la energía sería inagotable y libre, y supuso que pronto el clima podría controlarse, es decir que no era ningún pesimista cercano a la idea del *fin de la ciencia*, no acertase a prever esos usos tan decisivos de la máquina que tanto le debe. Von Neumann, en particular, fue siempre consciente de que lo que hacían al trabajar en el EPC, los dispositivos de cálculo con los que estaban ayudando a diseñar las armas nucleares en Los Álamos, no era simplemente mejorar en tamaño, cantidad y rapidez los enormes sistemas de que ya disponían. Según cuenta Dyson (2012: 222), Von Neumann le escribió en una carta a Lewis Strauss, un almirante que era decisivo para liberar los fondos requeridos para ejecutar su plan, que

El dispositivo proyectado, o, más bien, la clase de dispositivos de la que ha de ser el primer representante, es tan radicalmente nuevo que muchos de sus usos resultarán evidentes solo después de que haya sido puesto en funcionamiento. Esos usos que hoy no son predecibles, o que no lo son fácilmente, probablemente van a ser los más importantes,

es decir, que Von Neumann era plenamente consciente de que no estaba llegando a los límites de una idea, sino dando los primeros pasos en un panorama muy abierto, pero, al menos en ese momento, estrictamente limitado al cálculo, puesto que, un poco más adelante y en esa misma carta, Von Neumann se refiere únicamente a siete ámbitos de aplicación (Aerodinámica, Hidrodinámica, Elasticidad, Óptica, Electrodinámica, Meteorología y Clasificación) en los que el manejo de magnitudes cuantificables a velocidades enormes podría ser clave⁴.

⁴ El texto completo de la carta, de 24 de octubre de 1945, se conserva en el archivo dedicado a Von Neumann en la Biblioteca del Congreso de los EEUU. En una intervención de la hija de Von Neumann en una reunión de la John Von Neuman Computer Society se proporciona información muy inmediata sobre lo que el matemático pensaba estar exactamente haciendo. Accesible en <https://njszt.hu/en/neumann/hir/20130701/the-past-of-the-future-conference-%E2%80%93-an-account>.

Al tiempo que Von Neumann trabajaba con modelos nuevos y aparatos reales, Turing, además de ser un autor esencial en la especialidad, *filosofaba* sobre el caso, se preguntaba si las máquinas podrían pensar y sobre cuáles habrían de ser las condiciones en las cuales debiéramos atribuirles alguna especie de pensamiento, como lo muestra el que su artículo más influyente sobre el particular se publicase en la revista *Mind*. Sus especulaciones, tan interesantes como atrevidas y algo desordenadas, estaban muy ligadas al ambiente intelectual de su época, pero formulaban preguntas e ideales ambiciosos que muy probablemente hayan podido tener influencia en muchas de las bravuconadas que se han dicho al respecto por autores poseídos por la certeza de que estaban entrando en una era de descubrimientos y avances espectaculares, y que fueron algo menos circunspectos que el inglés. El hecho es que se hicieron toda clase de arriesgadas predicciones sobre la forma y la velocidad con la que las máquinas superarían en nuestras capacidades cognitivas, olvidando completamente la peculiarísima cualidad de lo que llamamos conciencia que es, desde luego, algo rotundamente distinto a cualquier programa.

Sin embargo, el impulso determinante para favorecer la generalización popular de esta clase de suposiciones no se debe exclusivamente a la filosofía seminal de Turing, sino que se dio, sobre todo, cuando los ordenadores *aprendieron a hablar, a leer y a escribir*, cuando dejaron de ser meros instrumentos de cálculo, una absoluta novedad, que, como se ha dicho, no fue prevista por los creadores de la informática, aunque si estuviera desde el principio en el imaginario de los pioneros de la IA. En el ámbito teórico, lo que ha facilitado muy mucho la suposición de que el paso de las máquinas a las personas no tendría que considerarse como algo irreal ni inalcanzable, es el hecho de que, como subraya Gelernter (1994: 150-159), el llamado *test de Turing*, debiera haberse interpretado como una prueba de que el ordenador es extremadamente inteligente, tanto o más que un humano, mientras que se ha entendido como una demostración de que simular y ser inteligente es exactamente lo mismo que simular y ser humano, de forma que la

prueba ha dejado de ser vista como una muestra de la inteligencia para convertirse en una prueba de la humanidad de la máquina. De ahí a suponer que, puesto que en el cerebro no existe ningún observador imaginario, no hay que exigir ninguna cualidad distinta a una máquina que su capacidad de hacer lo que hacemos para identificarla con nosotros mismos. Ese planteamiento olvida el problema que plantea la existencia de la conciencia, el hecho de que no se deba inferir que no existe porque no se pueda encontrar para ella un asiento corporal o espacial definido, y, en consecuencia, la deducción de que una máquina pueda tener exactamente todo aquello que nosotros suponemos tener, pero, como dice el propio Gelernter (1994:162), “*We might ultimately build a computer that seems to us to have a mind. But I doubt whether the computer itself will ever be taken in*”.

En cualquier caso, los seres humanos no somos puras mentes, tenemos un cuerpo sin el que el cerebro seguramente sería incapaz de hacer nada. Es decir que el paso de suponer que las máquinas tengan una mente a suponer que cuerpos y máquinas puedan fusionarse de modo que a través de ese cibernético rueden las ideas, los pensamientos y las emociones es un salto todavía más complejo. Al fin y al cabo, hasta Dennett (1991: 6) reconoce que la tecnología está todavía muy lejos de poder imaginar algo parecido,

Making a real but counterfeit coin is child's play, making a simulated coin out of nothing but organized nerve stimulations is beyond human technology now and probably forever.

2. Mentes, cuerpos y máquinas

Para no confundirse en esta clase de asuntos es conveniente no perder nunca de vista las diferencias entre la sintaxis y la semántica, y su muy distinto significado epistémico. Ahora tendemos a desentendernos de tal distinción porque nos hemos acostumbrado a que las tecnologías digitales se vean, cuando menos, como manifestaciones del conocimiento, incluso a que se lleguen a identificar con el conocimiento mismo.

Cabe pensar que sea un error tomarse en serio esta distinción, es lo que hacen quienes confunden sistemáticamente información (forma) con conocimiento (comprensión, algo ajeno a la forma por mucho que la reconozca, pues puede reconocer también lo informe), es decir asumir que la distinción equivale a dar por resuelta la cuestión. Podemos comparar esta fusión/indistinción con el famoso análisis ryleano sobre la *confusión categorial* que pretende que el error moderno en cuanto a la naturaleza del alma, lo que en otras ocasiones se ha llamado la *doctrina oficial*, o, en frase del propio Ryle, *el mito de Descartes*, “no es la comprobación de diferencias inexistentes, sino la falsa representación de diferencias existentes”⁵.

Pues bien, si retorremos un poco el argumento, podemos decir que quienes profesan la identidad de la mente con una máquina virtual, tomando al pie de la letra una metáfora que a Gelernter (1994: 122) le parece tan fascinante como confusa y falsa, admiten la identificación de unas diferencias insalvables, sencillamente porque creen no científica la manera de representarlas. Y claro es que no es científica, sino previa a cualquier ciencia, porque hacer ciencia es tratar de comprender algo real, en ningún caso crear un lenguaje que nos haga creer que comprendemos lo que no está, todavía, al menos, a nuestro alcance, y al categorizar la distinción intuitiva entre mental y físico, una diferencia sin la cual ni siquiera existirían tales palabras, y por compleja que resulte la separación de algunos de sus caracteres, no podemos atribuir la intencionalidad a nada distinto de una conciencia por muy reposada que cada conciencia esté en algo físico, mejor dicho, en algo vivo. Por otra parte, es necesario recordar que esa clase de maniobra se libra, por decirlo así, de la necesidad de entender la neurobiología, pues la analogía serviría igual si en lugar de cerebro, tuviésemos una calabaza, lo que es absurdo, pero coherente, en efecto, con la distinción entre el *software* y los dispositivos que lo ejecutan, que son como son, pero muy bien podrían estar hechos de melaza y cartón.

5 En Laslett 1957: 100.

Bertrand Russell, siendo él mismo sostenedor de muy paradójicas ideas al respecto, objetó fuertemente el análisis ryleano, rechazó que Ryle pretendiera haber resuelto una supuesta confusión categorial en el concepto de *mente* usando una mixtura peculiar de conductismo psicológico (aunque, que yo recuerde, sin citar a Turing) y de análisis del lenguaje ordinario, una escuela que Broad había motejado como la moda intelectual más extraña que podría existir. Russell pensaba que esta estrategia meramente verbal para analizar el problema de la mente implicaba posiciones inaceptablemente cerradas sobre cuestiones que era necesario plantear, aun asumiendo que no se conocerían las respuestas, pero le parecía muy pretencioso suponer que podrían despacharse como si se tratase de un equívoco. Pese a ello, en los últimos ochenta años se ha hecho común, no unánimemente, por supuesto, dar por hecho que la respuesta al problema que Russell planteaba, y Ryle pretendía *discover*, habría de consistir en que, por decirlo con palabras del propio Russell, un “físico idealmente competente” podría calcular, o predecir, “lo que Fulano de Tal diría el resto de su vida” (Russell 1976: 262).

Tal asunción guarda, de algún modo, un innegable aire de familia con cualquier pregunta acerca de si resultaría posible “transferir la mente humana a un soporte no biológico”. Al dar por hecho que las tecnologías digitales suponen conocimiento, su fantástica capacidad de proceso de datos nos invita a ver en sus resultados una especie de *pensamiento* similar al nuestro, nos invita a imaginar, de modo bastante ingenuo, que han *comprendido* información en el sentido semántico, que es como, de una manera resumida, podríamos denominar la actividad de una conciencia que se tiene por distinta de lo que contempla (y esta es una distinción real, lógica y discursiva) y que se hace preguntas basadas, ante todo, en la diferencia entre aquello que ahora examina atentamente, su experiencia pasada, y su mirada actual sobre el mundo, es decir que nuestra inteligencia está de hecho unida a una cualidad o dimensión consciente por la que se sale siempre del programa, es algo más que analítica y lógica. En los albores de la era digital, Stanislaw Ulam, uno de los matemáticos que trabajaban en Los Álamos, le

dijo a Von Neumann, “*What makes you so sure that mathematical logic corresponds to the way we think?*”⁶, Ulam quería señalar que ni cabe dar por supuesto que la mente sea una máquina virtual, ni tampoco que nuestros cerebros sean procesadores de tales máquinas, como no sea de una manera notoriamente imprecisa, puesto que los cerebros hacen muchas cosas que les están lógicamente prohibidas a los procesadores, un tipo de máquina que tiene unas características lógicas muy específicas y bastante rígidamente definidas, lo que permite, precisamente, que, al menos en principio, sea perfectamente realizable con cualquier clase de materiales con tal de que se atengan precisamente a la correspondiente receta.

Aunque se consigan sistemas de IA que puedan tener, como se viene diciendo que lo tiene *AlphaZero*, perspicacia y creatividad, esa cualidad tampoco puede confundirse con la conciencia, por mucho que, hasta la fecha, la hayamos visto habitualmente como su compañera, aunque no siempre, como es obvio, pues las formas de conciencia pueden ser muy variadas, y algunas bastante obtusas. Dicho de otra manera, es la inteligencia semántica la que crea la sintaxis, porque en el mundo físico no hay sintaxis alguna, con la posible excepción que plantea la idea de *objetividad*, y en especial la de los entes matemáticos, simplemente hay relaciones que se descubren y pueden formularse sintácticamente en la medida en que pueden ser descritas mediante un discurso. Pues bien, la digitalización es un recurso técnico, como la escritura, que se emplea para cifrar un signo que es una creación semántica que relaciona tres categorías muy distintas, realidad, percepción y lenguaje, para crear una realidad física que soporta intenciones, propiedades semánticas.

3. Máquinas y seres vivos

Máquinas y seres vivos se diferencian de muchas maneras. La capacidad de autorreplicación, la reproducción, todo lo que hace que la vida esté poblada por *sujetos*, por auténticos individuos que se relacionan, iguales pero

⁶ Tomado de https://www.edge.org/conversation/george_dyson-turings-cathedral.

muy distintos, de modo que esa disparidad básica explica que las posibilidades de unas y otros sean muy otras. Hay una notable comunidad entre los seres vivos, algo que tiene que ver con cierta unidad, eso que ahora se conoce como ecología, que, como tal, para nada aparece en los seres inanimados, aunque ahora podamos pensar en una suerte de imitación de esas relaciones mediante lo que se ha llamado *Internet de las cosas*. Otra propiedad exclusiva de los seres vivos es su capacidad de alimentarse, crecer y transformarse, funciones que pueden ser imitadas, pero que en la vida tienen un estatuto fundamental que todavía no sabemos cómo explicar bien. También es importante tener en cuenta la peculiar forma en la que se relacionan los diferentes componentes o elementos de los seres vivos mediante un tipo de interacciones de todos con todos que no son ni siquiera fáciles de representar, y aunque sea evidente que llamar a esto *diseño inteligente* no resuelve ningún problema, tampoco parece muy prometedor dar por hecho que sea trivial llegar a comprender las leyes que rigen esa especie de comportamientos.

Si nos fijamos en el modo en que trabaja, es evidente que nuestro cerebro está muy por delante en versatilidad, en reconocimiento de formas y diferencias, y en capacidad de adaptación respecto a los sistemas de computación, que supera ampliamente muchos de sus rendimientos, a excepción de tareas muy especializadas como recuperación de información o cálculos, además de ser mucho más parco consumiendo energía que cualquier máquina. Es verdad que podemos esquematizar algunas de sus funciones, pero sin perder nunca de vista que imitar los resultados no supone conocer la forma en que el cerebro los obtiene. Si lo supiésemos, no tendríamos que obligar a un sistema de reconocimiento visual a estudiar decenas de miles de imágenes hasta llegar a saber lo que es un plátano, por ejemplo, cuando un niño puede aprenderlo con un par de vistazos.

No deberíamos olvidar que las tecnologías son una aplicación de propiedades que somos capaces de discernir y aprovechar en la realidad, que nuestros *cerebros electrónicos* son criaturas fabricadas para hacer mejor con ellas lo que nos resulta muy difícil hacer sin ellas, es decir que pertenecen

a la saga de los martillos, las ruedas y las palancas, entidades que pueden verse como una estilización de cualidades naturales que ha sido creada por seres inteligentes que han *visto* la manera de potenciar y utilizar en su provecho y para su invención determinadas posibilidades, relaciones o leyes físicas.

Lo que nos puede confundir es que ahora poseamos máquinas que manejan símbolos y no solo cosas, una posibilidad que se empezó a hacer factible a partir de una ocurrencia de Von Neumann ciertamente revolucionaria, por gastada que pueda considerarse esta expresión, a saber, que los números que manejaban los dispositivos de cálculo que estaba tratando de crear se podrían emplear de dos maneras muy distintas, tal como lo ha subrayado George Dyson⁷

By breaking the distinction between numbers that mean things and numbers that do things, von Neumann unleashed the power of the stored-program computer, and our universe would never be the same.

A partir de ese cambio fue posible empezar a manipular símbolos mediante sistemas físicos (es lo que llamamos *software*), es decir a entendernos con máquinas digitales a través de programas escritos en símbolos, lo que ha permitido un desarrollo casi inimaginable de posibilidades que pueden llegar, incluso, a descubrir propiedades y relaciones que se nos hayan escapado a primera vista, especialmente cuando se trabaja sobre bases de datos exhaustivas manejando procesos de cálculo inabarcables para la capacidad del cerebro humano que, a este respecto, puede ser descrito como una *máquina* cuyo *algoritmo* nos es desconocido.

Como hemos observado a propósito del éxito de *AlphaZero*, nuestros algoritmos podrían llegar a ofrecernos resultados cuyo intrínquilis se nos escape, es decir, podrían producir una *inteligencia sin conciencia*, cuyos resultados pudiéramos verificar, pero no siempre comprender, una perspectiva que

⁷ Tomado de https://www.edge.org/conversation/george_dyson-turings-cathedral

sirve, a mi entender, para aclarar más que para minimizar las variadas diferencias insalvables que ahora mismo separan cualquier máquina de un ser vivo y consciente.

Por esa razón, no habría que confundir la cuestión de si podremos seguir fabricando exoórganos, con la de si podremos desprendernos de nuestro propio cuerpo para instalarnos en otro más duradero. La primera cuestión admite, evidentemente una respuesta afirmativa, puesto que lo que hacemos es prolongar con nuestra inteligencia y su capacidad inventiva las capacidades sensoriales y motoras de nuestros órganos, de manos, oídos y ojos; en el segundo caso, de ser posible, haríamos algo equiparable a abandonar nuestra sombra, una posibilidad que no es solo fantástica, sino que comete el abuso de confundir al actor con su obra. Los sistemas híbridos pueden llegar a ser muy eficaces, como un piloto con su avión o una *influencer* con su *smartphone*, porque esos instrumentos pueden llegar a prestar servicios inimaginables hace sólo unos pocos años, tales como hacer traducciones simultáneas o pagar un platero comprado en Guanajuato cuando estamos al otro lado del Atlántico, pero no son *nosotros*, como tampoco lo son por sí mismas nuestras manos.

Dentro del cuerpo, o fuera de él, las tecnologías de *mejoramiento humano* podrán, seguramente, ampliar nuestra esperanza de vida y hacer menos penosos el envejecimiento y la enfermedad, pero no podrán proporcionar respuestas a preguntas que nosotros no sepamos formular o cuya clave no les hayamos infundido. Puede que lleguemos a poseer un conjunto extraordinariamente atractivo y poderoso de nuevas tecnologías habilitantes, pero nada de eso nos dirá algo interesante acerca del sentido y el valor de la vida, de la personal y de la colectiva, ni existe la menor base para suponer que podamos trasladar cualidades de la vida a soportes artificiales, que podamos hacer conscientes a realidades que carecen de las extraordinarias cualidades de lo vivo, y que, aunque pudieran llegar a ser más capaces de entender y resolver problemas que nosotros no lo serían de manera consciente, subjetiva.

Actualmente, la arquitectura de las computadoras se basa en una unidad de procesamiento central (CPU) que controla y/o se auxilia de unidades que se especializan en tareas específicos, como las unidades de procesamiento gráfico (GPU), y es perfectamente imaginable que en algún momento se pueda conectar un sistema informático con lo que se puede llamar una NPU, una unidad de procesamiento neuronal que funcionase tomando directamente de un cerebro los datos o señales que ese cerebro maneja (claro es que en la medida en que se puedan aislar esos procesos, comprender sus fines, y aprender a interactuar con ellos sin destruirlos) permitiendo que la CPU pudiese contribuir de manera significativa a hacer cosas que el cerebro por sí solo no podría hacer, pero esto no sería nada muy distinto de lo que ya podemos hacer cuando un exoesqueleto artificial ayuda a andar a una persona que tenga dificultades para mover sus piernas.

Podemos tratar estas cuestiones acudiendo a la diferencia entre partes del cerebro, con las que podríamos interactuar como interactuamos con manos y pies, y el *cerebro completo* (con sus terminales en todo el cuerpo y su capacidad de percibir y razonar) en cuanto órgano y asiento de la mente consciente, que está bastante más allá de nuestra comprensión y cuya sustitución o fusión plantea cuestiones que no son meramente orgánicas o instrumentales, aunque solo sea por el hecho de que ignoramos lo esencial de cuanto se pueda preguntar acerca de qué y cómo hace el cuerpo para tener conciencia, y de qué y cómo hace el cerebro para ser inteligente. Estas cuestiones no se podrán encomendar a nadie distinto de nosotros mismos porque podremos ampliar nuestro cuerpo, pero sin salirnos de él para contemplarlo desde fuera, ni para dejar de serlo en la forma en que lo somos.

Cuando se excede la imaginación y pensamos mediante una analogía insuficiente que lo mismo que se *prolonga* nuestro cuerpo podremos fusionar nuestra mente con sistemas que la doten de propiedades transformadoras, se comete, sin duda, un abuso. Tras esta clase de pretendidas analogías, se oculta una confusa metafísica, que la tecnología no analiza porque no puede utilizar. Y no puede utilizarla porque la tecnología se mueve en los límites del reino de la realidad y el de lo posible, un ámbito al que inicialmente

accedemos mediante cierta especie de fantasía y eclipsando aquellas cuestiones para las que no tenemos respuesta suficiente. Lo corriente es negar una metafísica explícita, o pretender que el materialismo no lo es, que es un retrato esencial de la realidad, pero por mucho esfuerzo que se haga en huir de afirmaciones acerca de esas cuestiones, su relevancia es innegable y existe lo mismo si se formula que si se pretende eclipsar.

Cabe decir que muchos *profetas* tecnológicos abusan de nuestra credulidad mediante tres procedimientos distintos: en primer lugar, dar la impresión de entender lo que realmente no entienden, como, a propósito de un tema distinto, ha dicho Ethan Siegel⁸

No amount of fancy probabilistic analysis can justify treating guesswork and wishful thinking as having any sort of scientific weight. Applying scientific techniques to an inherently unscientific endeavor, such as inventing estimates to unknowns about the Universe, doesn't make it any more scientific. The opposite of knowledge isn't ignorance; it's the illusion of knowledge.

En segundo término, confundir con verdades indiscutibles lo que apenas son metodologías que ya se verá lo que consiguen, y, finalmente, olvidar la inmensa cantidad de cosas que se ignoran acerca del funcionamiento básico de la vida, desde cómo se mueve un corazón, hasta cómo se realiza la síntesis clorofílica, o qué hacer para parar los pies a un virus asesino cuyo despiece a nivel molecular es suficientemente conocido. Saltarse la vida, puede ser audaz, pero no parece muy razonable pretender el logro de éxitos trascendentales haciendo mejor que ella cosas que claramente no comprendemos cómo hace.

4. La realidad irreductible y lo artificial

La ocultación de las implicaciones metafísicas, o su liquidación mediante argucias verbales e ideas elusivas, dificulta la comprensión de

⁸ Tomado de <https://medium.com/starts-with-a-bang/no-we-havent-solved-the-drake-equation-the-fermi-paradox-or-whether-humans-are-alone-8f31a559f741>

preguntas que ponen en un serio aprieto nuestra capacidad de comprender. Me refiero a cuestiones que cualquiera puede entender, pero que no admiten respuesta definitiva, al menos hasta ahora y que aparecen supuestamente resueltas en las propuestas proféticas de los transhumanistas o tecnohumanistas más exaltados, de aquellos que suponen que podremos depositar nuestra mente (ya que *es* nuestro cerebro) en un soporte informático; son dimensiones de la realidad que, si se analizan con un mínimo de atención, marcan unos límites bastante definidos a la verosimilitud de cualquiera clase de fantasías hipertecnológicas. Me fijaré, en particular, en cinco grandes conceptos, *existencia, vida, mente consciente, objetividad y libertad* a los que dedicaré un brevísimo análisis en relación con el tema que nos ocupa.

El término *existencia* es, como se sabe, un tecnicismo filosófico que ha adquirido múltiples resonancias, pero ahora puede servirnos para recordar algo que se olvida en las propuestas de trasladar nuestra mente-cerebro a un soporte informático. Resulta que existen cosas, que hay una realidad y que nada ni nadie está en condiciones de crearla, que todo lo que podemos hacer es, primero, contemplarla, que es lo que hace la ciencia, y usarla o transformarla, que es lo que hace la vida práctica y la tecnología. Pero la contemplación y el uso presuponen la existencia y afirman nítidamente su irreprimible independencia, esa es la primera lección que aprendemos al reflexionar correctamente, que las cosas son lo que son sin que importe poco o mucho lo que pensemos de ellas pues, como dijo Philip K. Dick, la realidad es aquello que continúa existiendo y no desaparece cuando dejamos de creer en ella. La realidad es inagotable para nosotros, y no sabemos por qué es como es y no de otra forma. ¿Por qué hay la realidad que hay y porqué es de la manera que es? Es la pregunta que surge, y que no puede contestarse como si realmente fuéramos capaces de crear un marco en el que dar respuesta a una cuestión tan general y básica, pero no debiera resultar difícil reconocer que no somos capaces de tal hazaña.

La vida es aquella parte de la realidad natural que más nos sorprende, que nos parece más rica y compleja. No posee la grandiosidad de los espacios infinitos ni la energía de los acontecimientos cósmicos que podemos contemplar, suponemos o imaginamos, pero es una realidad tan extremadamente peculiar y atractiva que nos desborda por todas partes y en la que centramos la mayor parte de las energías que dedicamos a comprender el mundo. ¿Cuál es el problema aquí?: que tratemos de comprender la vida a partir de lo que ella no es, que nos limitemos a analizarla por sus costuras, a reducirla a los elementos con que cuenta, como si fuese un mero constructo de lo que de suyo no tiene vida, dando por hecho que nuestra metodología científica, acercarnos paso a paso a lo más complejo desde lo más simple, nos permitiera reducir la realidad de la vida a otras esferas de la realidad más simples y manejables, lo que, en último término, supone enmascarar la ausencia de investigación científica consistente, es decir, de vías precisas de explicación sometible a contraste, con mera palabrería. “Complejidad” es una palabra hermosa, pero de ella cabe decir lo que Francis Bacon anotó sobre la esperanza, que es un buen desayuno, pero una mala cena.

Es interesante llamar la atención sobre el hecho de que la Biología como ciencia profese una cierta tendencia a renegar de la vida, a decir que “la vida no existe”, como hizo Kahane⁹, o a resignarse a reconocer a la muerte, y no solo a la humana, un verdadero significado biológico como puso de manifiesto J. D. Vincent (2001)¹⁰, una situación sobre la que llamó la atención Feynman en el *Galileo Symposium* de septiembre de 1964, cuando dijo que “*there is nothing in biology yet found that indicates the inevitability of death*”, aunque, con su característico optimismo, posiblemente un tanto irónico en esa ocasión, sugería que tal vez eso significase que ese magno problema pudiese ser resuelto si la Biología diese alguna vez con el *quid* de la cuestión, pero el

⁹ Citado por Jean Rostand, 1970.

¹⁰ La cita corresponde a su intervención “What Is and What Should Be the Role of Scientific Culture in Modern Society”. La conferencia está incluida en Feynman (2000: 85).

hecho es que los biólogos no parecen demasiado preocupados por tal *problema*, ni siquiera cuando se lanzan a filosofar¹¹.

Detrás de esta clase de prejuicios podemos encontrar una vieja obsesión metafísica, la pretensión de que la realidad debe ser una y no plural, la convicción casi instintiva de que del mismo modo que nuestra mejor manera de explicar algo consiste, aunque no siempre, en un proceso reductivo, en la detección de lo primitivo, lo indivisible, lo atómico, la realidad debe estar formada a partir de una única estofa, nunca de dos, y, menos aún, de más. Se trata de una presunción que no busca razones porque nuestro entendimiento tiende autónomamente a darle razón, a sustancializar su propia senda de búsqueda, a convertir su acto en objeto. Como ha escrito Juan Arana (2005: 244),

Quando cierto concepto nos ha permitido salir airosos de una serie de lances, hay que tener mucha contención para darse cuenta de que sigue siendo un concepto y no la realidad realmente real.

Por aquí llegamos a la tercera gran cuestión, la central, sin duda, en este tipo de análisis. Cuando se admite que nuestra mente, es un programa que corre en el ordenador de nuestro cerebro, ni siquiera se dice que sea algo que “podría ser imitado por un programa”, porque esa expresión pone de manifiesto la dificultad del caso, de algo que no tenemos, mientras que la afirmación desnuda y rotunda deja al lector ingenuo con la impresión de que seguramente se ha perdido algo, que quién es él para dudar de algo que aseguran tantos sabios. ¿De dónde ha salido, pues, la curiosa idea de que algo tan singular como una mente consciente podría depositarse en un soporte distinto al que la realidad natural ha dispuesto para ese modo tan peculiar de existencia?

La posibilidad misma de esa afirmación se funda en tramos de nuestra historia intelectual que es necesario recordar. En primer lugar, el hecho de

¹¹ Por ejemplo, en el índice temático del volumen sobre filosofía de la Biología editado por Ayala y Dobzhansky (1983), ni siquiera figura el término *muerte*.

que una buena mayoría de cultivadores de la filosofía de la mente hayan asumido que el problema fundamental era el de la objetividad de los *qualia*, y que si la ciencia los había colocado en una especie de limbo, habría que procurar su identificación con los procesos cerebrales, que es el asunto del que se ocuparon básicamente Ullin Place (1956) y Jack Smart (1959), de forma que creyeron que esa tarea era fácil de llevar a cabo, dando lugar a que, como más tarde subrayó Jaegwon Kim (1971: 336),

...mientras que, para Platón, los pensamientos racionales eran las actividades paradigmáticas del alma, los hechos tan menudos como picores, dolores de muelas y punzadas están ahora entre los típicamente mencionados en la discusión filosófica como eventos mentales,

...una situación ciertamente curiosa para cualquier mediano conocedor de la historia previa de esta cuestión.

De manera paralela, quiero decir, lógicamente independiente, el funcionalismo se tomó en serio la idea de que la actividad mental podría ser identificada con el proceso lógico de símbolos, pasando de un análisis procesual a una identificación de carácter ontológico, dando por sentado que el carácter decisivo de los soportes de cualquier tipo de conciencia era su condición funcional, no su carácter material, una idea tan audaz como realmente improbable que a los biólogos siempre les ha producido una desazón fundamental, pues como escribió Crick (1994: 94), “es tan estafalaria que hay muchos científicos que se asombran al saber que de verdad existe”; más displicente, si cabe, resulta la respuesta de Freeman Dyson a la pregunta de Edge sobre si las máquinas piensan (Brockman 2017: 79):

I do not believe that machines that think exist, or that they are likely to exist in the foreseeable future. If I am wrong, as I often am, any thoughts I might have about the question are irrelevant. If I am right, then the whole question is irrelevant.

Tampoco deja de ser sorprendente, dicho sea de paso, que la posición que goza de mayor favor entre los pensadores que se sienten más cercanos

a la Biología sea el emergentismo, que, de alguna forma, se podría caracterizar como una especie de retroceso conceptual hacia *explicaciones* similares a las de la *generación espontánea* que Pasteur parecía haber liquidado definitivamente.

Estas dos líneas de ataque a la supuesta fortaleza ontológica del *cartesianismo*, un *ismo* que permite olvidar la sutileza del francés y confunde su *solución* con el problema que le mueve a pensar, se han concedido sin ninguna falsa humildad la respectiva victoria, así es que cualquiera que quiera reexaminar el asunto se tropezará con una escolástica contraria y suficientemente arisca. En ambos casos se olvida el carácter irreductiblemente singular de cada conciencia, la paradoja fundamental de que nadie haya experimentado jamás otra que la propia, circunstancia nada irrelevante en la que se apoyó el conductismo psicológico que negaba la utilidad de la noción de conciencia partiendo de las dificultades para objetivar sus cualidades y/o actos, y así procedió a edificar su castillo de naipes: como no podemos conquistar ni medir la montaña, la montaña no existe.

La cuarta consideración metafísica con la que querría mostrar la restricción de la realidad que se asume en los proyectos reduccionistas y funcionalistas es la que tiene que ver con la sorprendente existencia de la objetividad, con el hecho de que seamos capaces de conocer verdades más allá de toda duda, con nuestra asombrosa capacidad para mostrar algo tan extraordinariamente elegante y admirable como la demostración de que el número de los números primos es infinito, cosa que se puede hacer con toda facilidad gracias al genio de Euclides. En más de una ocasión he insistido, apoyándome en una lectura de Hermann Weyl (1952), que éste es esencialmente el problema que se le planteó al Descartes físico y filósofo, tratar de entender cómo la necesaria libertad de la mente consciente, que excluye la posibilidad de cualquier coerción física a la hora de sentir y razonar, resulta un problema insuperable para cualquier proyecto de naturalización de la conciencia, una constatación indiscutible que Descartes cifró con el *cogito ergo sum*, y cómo ese descubrimiento ineludible que exige una *libertad negativa* respecto a los límites del cuerpo, sea compatible, de manera un tanto

paradójica, con el que perdamos cualquier clase de *libertad* al hacernos conscientes de cualquier evidencia inesquivable, y, en especial, ante esa primera certeza en su forma más radical que nace de la síntesis de una evidencia lógica con la sensación de existir, con la constatación de que, por decirlo de manera provocativa, en este punto decisivo, *esse est percipi*.

El hecho de que podamos simbolizar la forma de nuestra certeza en proposiciones lógicas, el que podamos convertir evidencias en símbolos, es el que ha llevado a confundir la capacidad de pensar, que es una función esencial de la vida en muchas de sus formas, con la capacidad de manipular símbolos y ese error categorial es el que se comete una y otra vez cuando se quiere convertir a la IA en heraldo de una metafísica *científica y tecnológica*.

Habitualmente se suele entender por inteligencia, entre una gran variedad de funciones, la capacidad para reconocer el entorno, establecer relaciones formales y causales, perseguir propósitos y evaluar las dificultades que se oponen a ellos para tratar de superarlas. Tal conducta implica normalmente lo que conocemos como conciencia y/o pensamiento, pero no es contradictorio que esas capacidades puedan alcanzarse en forma algorítmica y en contextos restringidos, sin la presencia de ellos, una inteligencia eficaz, que, a su manera, pensase (pues pensar es pesar, calcular, comparar, etc.) pero no de manera consciente.

La IA no busca únicamente máquinas, o sistemas informáticos, capaces de resolver problemas muy difíciles, sino que, con intención más ambiciosa, busca que sus sistemas lleguen a poseer las capacidades cognitivas que se tienen como características de los seres humanos, tales como aprender, resolver problemas, razonar mediante analogías, escoger ejemplos del pasado a través de la memoria, etc., pero confundir esa búsqueda con una forma de hacer que se evapore completamente la cuestión de la conciencia o, de la inteligencia consciente, y hasta sentiente, si se quiere, sería una forma de hacer trucos con el lenguaje.

Convertir a la IA en heraldo de una metafísica científica, solo puede hacerse partiendo de que se ha identificado, previamente, la mente con la conciencia, la conciencia con la inteligencia, y la inteligencia con lo que

constituye su objeto, o de otro modo, su resultado, olvidando completamente su carácter intencional, su subjetividad inquisitiva, fijándose únicamente en aquello que obtiene. Cualquier actuación de la IA es, en último término, consecuencia de una inteligencia humana previa, y puede lograr resultados portentosos porque, no lo olvidemos, pueden ser ejecutados ya que las fases que conducen a ellos han sido exhaustivamente descritas y posteriormente incorporadas en una serie de instrucciones de funcionamiento, aunque, como hemos dicho, no haya que descartar del todo que algoritmos poderosos y muy bien diseñados y dirigidos puedan *aprender*, mejorar sus métodos y alcanzar resultados que puedan desbordar, de algún modo, las previsiones iniciales, que puedan llegar a descubrir formas de inteligibilidad que nos asombren y que solo podamos verificar como correctas en su aplicación a casos, es decir, que podamos confiar plenamente en ellas y usarlas a diversos propósitos sin entender suficientemente cómo funcionan.

En resumen, ¿cuál es la verdadera dificultad para no ceder a las pretensiones de quienes creen que una mente podría ser objetivada y, en último término, conectada a sistemas artificiales que le prestasen una serie de cualidades, por ejemplo, una perdurabilidad que el cerebro biológico no posee? Pues que la conciencia no ha sido descrita ni, mucho menos, formalizada, porque, hablando con propiedad, no puede serlo, ya que su ser es percibir, y su percibir es estar siendo, es temporal, fugaz, está viva y cambia conforme a pautas que no se pueden dominar porque pertenecen a la misma esencia de la identificación de la conciencia con el tiempo que pasa, que no es el tiempo del reloj, sino el de la existencia.

Para acabar, quisiera apuntar un par de cuestiones relacionadas con el último panorama metafísico al que me he referido, con la libertad intelectual y moral, para mostrar como ese panorama futuro de trasplantes de conciencias a máquinas constituiría un mundo absolutamente absurdo, bastante más que éste, sin duda. La *inmortalidad tecnológica* no sería, por cierto, la menor de sus desventajas.

El concepto de libertad es, probablemente, el más problemático que existe, nada es tan fácil como producir equívocos al respecto. Esta dificultad tiene, a mi modo de ver, una raíz esencial en el hecho de que la libertad, como la existencia o la conciencia que son nombres que acuden con facilidad a nuestra mente cuando tratamos de entender asuntos de este tipo, son conceptos elusivos, que no ceden fácilmente a la cosificación. Muy en especial, la libertad de acción acaba por ser incompatible con cualquier clase de explicación, puesto que una vez que cualquier acción resulta explicada aparece, *eo ipso*, como necesaria, la libertad se evapora, del mismo modo, podríamos decir que, como subrayó Kant, cuando describimos suficientemente cualquier realidad no ganamos nada diciendo que esa realidad existe porque la mera existencia no se nos da nunca por modo de concepto.

En consecuencia, cuando queremos entender en que consiste la libertad, es frecuente que se la acabe reduciendo a su aspecto puramente práctico, a la manera de Hobbes (2000: 167)

La libertad, si quisiéramos definirla, no es otra cosa que una ausencia de obstáculos que impiden el movimiento. Así, el agua que está contenida en un vaso no tiene libertad porque el vaso mismo la impide salir; mas si el vaso se rompe, el agua queda liberada. De igual modo, el hombre tiene mayor o menor libertad, según tenga más o menos espacio en el que moverse; así, quien está encerrado en una prisión grande, goza de más libertad que quien lo está en una pequeña. Y un hombre puede que sea libre hacia un lado, y no lo sea hacia otro, así sucede con el viajero que está limitado por setos o vallas laterales, para que no estropee las viñas o sembrados de maíz que crecen a ambos lados del camino.

Cualquier concepto de libertad un poco más sofisticado, tropieza con dificultades bien conocidas. Es lo que pasa con las explicaciones de la conducta que, a medida que se aceptan no dejan ningún espacio a la libertad que se ve sustituida por diversas entidades como los *motivos*, el *carácter*, los *genes* o la *sociedad*. Cualquier libertad deja de existir en esas perspectivas: pues todo lo que hacemos depende de cómo somos, lo que somos es independiente de cualquier decisión personal, no existimos porque queramos

ni como queremos, y, en consecuencia, nadie puede responsabilizarnos en último término de nada de lo que hagamos, y esta forma de argumentación, que podría constituir un buen alegato a favor del perdón y la piedad, implica también una cierta asunción pesimista del determinismo más radical, de que alguna forma de predestinación es inevitable.

Una mentalidad de este tipo está, a mi modo de ver, por debajo de lo que suponen aquellos que propugnan la desaparición de la humanidad, la muerte de todos para siempre, lo que no es poca cosa, y creen que un mundo de mentes artificiales tendría menos defectos que el nuestro, tal vez porque sería un mundo al que le habríamos privado radicalmente del sentimiento de culpa, incluso aunque hubiese acabado con nosotros. Espero que a nadie se le escape que semejante tipo de explicaciones son pura locura, ejemplos extremos de inconsecuencia, formas de idolatrar una Naturaleza de la que habríamos de ausentarnos para no interferir con sus divinales propósitos.

Una máquina nunca puede ser libre, su única, y paradójica, *libertad* se daría cuando no funcionase, cuando dejara de ser lo que es, y de poder hacer lo que podía hacer, para comportarse conforme a la estofa informe de la que está hecha. Nuestra supuesta máquina cerebral como soporte de la mente consciente tiene necesariamente que dejar un espacio a esa libertad intelectual que es necesaria para decidir con razón entre opciones, y que es la única capaz de llevarnos a reconocer la verdad, aunque ese reconocimiento, como ya queda dicho, produzca el efecto paradójico de que la libertad del intelecto tenga que aquietarse y reconocer lo que es inequívocamente verdadero.

Hace ya mucho que, en nuestra tradición cultural y moral, se asienta la idea de que solo la verdad nos puede hacer verdaderamente libres, y esa peculiarísima condición es la que cuadra a nuestra mente consciente, a nuestra capacidad intencional, y también a esa libertad radical que es la capacidad de actuar por exigencias morales que, como tales, solo pueden fundarse en nuestra voluntad de aceptarlas, de rendirnos incondicionalmente ante ellas.

Es evidente que esa realidad queda restringida cuando pensamos únicamente en un universo de esquemas, de máquinas, cuando se pretende que la tecnología no se limite a modificar la realidad, sino que sus logros se conviertan en la única clave interpretativa acerca de lo real. En esto consiste la estrechez metafísica de la tecnología, una limitación de horizonte que no está, obviamente en las tecnologías mismas, sino en las imágenes metafísicas del universo que absurdamente se pretenden *descubrir* y justificar con ellas.

Esta crítica no supone la negación de ninguna posibilidad humana interesante, es simplemente un recuerdo de lo fácil que puede resultar equivocarse. No hay duda de que los seres humanos seguirán produciendo novedades que siempre podrán parecer asombrosas a quienes las contemplen, y no digamos nada sobre lo que podrían parecer a los que murieron muy antes de verlas, lo que implica que el hombre que llegare a la conclusión que parodiaba el poeta¹²,

El hombre es por natura la bestia paradójica,
un animal absurdo que necesita lógica.
Creó de nada un mundo, y su obra terminada,
“ya estoy en el secreto —se dijo— todo es nada”

estaría incurriendo en un error muy de fondo que arruinaría su capacidad de pensar lógicamente, desproveería de cualquier valor a sus análisis, y, por supuesto, le haría completamente incapaz de invención, pues esa capacidad se basa esencialmente en la idea de que el mundo es más de lo que se ve, nunca menos, jamás nada.

Soy consciente de que todo lo dicho puede tomarse como literatura, y no especialmente buena, por cierto, aunque se trate de una literatura que busque precisión, pero pienso es el tipo de literatura que es necesario cultivar cuando se trata de hacer frente a los relatos que nos hablan del advenimiento inmediato de la *singularidad*, del *Mind up-loading* y lindezas

¹² Antonio Machado, *Campos de Castilla*, Proverbios y cantares, XIX.

similares. Se trata de recordar, en suma, que la realidad, esa peculiarísima suma de lo *exterior* y lo *interior* que es el mundo humano, aunque en ella esté nuestro medio y nuestro sostén, es ancha y ajena y que, como dijo Heráclito (Kirk y Raven, 1970: 235),

Por mucho que camines y lejos que llegues, no alcanzarás a encontrar los límites del alma: tan profundo es el logos que ella tiene.

Lo mejor que podríamos decir, por tanto, al que creyere haberla encontrado en un cachivache, es que, sencillamente, se habría confundido de mundo.

Bibliografía empleada

- J. Arana, *Los filósofos y la libertad*, Madrid, Síntesis, 2005.
- F. J. Ayala, Th. Dobzhansky, *Estudios sobre filosofía de la biología*, Barcelona, Ariel, 1985.
- J. Brockman, ed., *La tercera cultura. Más allá de la revolución científica*, Barcelona, Tusquets, 1996.
- J. Brockman, ed., *What to Think about Machines that Think*, New York, Harper, 2015.
- F. Crick, *La búsqueda científica del alma*, Madrid, Debate, 1994.
- D. Dennett, *Consciousness explained*, Little, Brown, Boston 1991.
- C. Diosdado, F. Rodríguez Valls, y J. Arana, eds., *Neurofilosofía. Perspectivas contemporáneas*, Thémata y Plaza y Valdés, Sevilla y Madrid, 2010
- G. Dyson, *Turing's Cathedral: The Origins of the Digital Universe*, New York, Vintage, 2012.
- R. Feynman, *El placer de descubrir*, Barcelona, Crítica, 2000.
- D. Gelernter, "El estudio del Talmud", en Brockman ed., 1996: 227-235.
- D. Gelernter, *The Muse in the Machine*, New York, Free Press, 1994.
- J. L. González Quirós, *Mente y cerebro*, Madrid, Iberediciones, 1994.
- J. L. González Quirós, *El porvenir de la razón en la era digital*, Madrid, Síntesis, 1998.
- J. L. González Quirós y J.L. Puerta, "Tecnología, demanda social y medicina del deseo", *Medicina Clínica*, Barcelona Elsevier-Doyma, 2009 (133(16)): 671-675.
- J. L. González Quirós "La teoría de la mente: de la inteligencia artificial a la inteligencia híbrida", en C. Diosdado, F. Rodríguez Valls, y J. Arana, eds., *Neurofilosofía. Perspectivas contemporáneas* 2010:153-176.
- Th. Hobbes, *De Cive* (1642), ed. Carlos Mellizo, Madrid, Alianza, 2000.
- J. Kim, "Materialism and the criteria of the mental", *Synthese*, 1971 (22): 323-345.

José Luis González Quirós

- G. S. Kirk, y J. E. Raven, *Los filósofos presocráticos*, Madrid, Gredos, 1970.
- R. Kurzweil, *The Age of Spiritual Machines*, New York, Penguin, 1999.
- R. Kurzweil, *The Singularity is Near*, New York, Penguin, 2005.
- P. Laslett, ed., *Las bases físicas de la mente*, Buenos Aires, Nueva visión, 1957.
- J. Pearl y D. Mackenzie, *The Book of Why? The New Science of Cause and Effect*, Londres, Penguin, 2018.
- U. T. Place, "Is consciousness a brain process?", *British Journal of Psychology*, 1956 (47:1): 44-50.
- J. Rostand, *El correo de un biólogo*, Madrid, Alianza, 1970.
- B. Russell, *La evolución de mi pensamiento filosófico*. Madrid, Alianza, 1976.
- J. Smart, "Sensations and Brain Processes", *Philosophical Review*, 1959 (68): 141-156.
- G. Strawson, *Things That Bother Me*, New York, New York Review of Books, 2018.
- S. Strogatz, *Infinite Powers: How Calculus Reveals the Secrets of the Universe*, Boston, Houghton Mifflin (publicación prevista para abril de 2019).
- A. M. Turing, "Computing Machinery and intelligence", *Mind* 1950, (59,236): 453-460.
- J. D. Vicent, L. Ferry, *Qué es el hombre*, Madrid, Taurus, 2001.
- H. Weyl, *Philosophy of Mathematics and Natural Science*, New York, Atheneum, 1952.

José Luis González Quirós
jlgonzalezquiros@gmail.com