

INTELIGENCIA ARTIFICIAL Y PENSAMIENTO HUMANO (Un enfoque raciovitalista de la discusión sobre la Inteligencia artificial)

J.M. Atencia López. Universidad de Málaga

Resumen: Nuestro propósito es realizar una aportación a la discusión sobre la Inteligencia Artificial desde la perspectiva raciovitalista. Examinamos las dificultades de la equiparación entre inteligencia artificial y pensamiento humano para destacar la especificidad de este último y su irreductibilidad al procesamiento de información del ordenador. Partiendo de la aceptación de las argumentaciones realizadas durante los últimos años por H. Dreyfus, pensamos que las tesis raciovitalistas sobre el hombre y el pensamiento suministran, una vez más, un punto de vista clarificador no sólo sobre las diferencias entre hombre y máquina, sino también sobre la identidad de lo h como tal.

Abstract: Our aim is to contribute to the discussion about Artificial Intelligence from a rational-vitalist approach. First we consider the difficulties to compare Artificial intelligence and human thinking to emphasize the uniqueness of the latter and its irreductibility to the computer data processing. Starting from the argumentations developed by H. Dreyfus in the last years, we think that once again the rational-vitalist theses about man and thinking provide a revealing point of view about the differences between men and machine and, what's else, about the identity of human nature.

1. Inteligencia artificial

Si tomamos 1956, fecha de la conferencia de Dartmouth, como el inicio de la andadura de la comunidad científica formada en torno a la Inteligencia artificial (IA), tenemos que reconocer que se han alcanzado (1) logros espectaculares, pero no el cumplimiento de la aseveración de A. Turing, para quien a finales de este siglo el uso de las palabras y la opinión educada común habrá cambiado tanto que uno podrá hablar de máquinas pensantes sin temer ser contradicho (2).

En su progreso se han podido distinguir cinco fases (3), correspondientes a otros tantos tipos de ordenador. La primera «generación», dotada de un *hardware* de tubos de vacío (ése era el caso de ENIAC), apareció en el mercado a principios de la década de los cincuenta y podía sumar dos números en menos de una milésima de segundo. La segunda, en la que los recién inventados transistores reemplazaron a los tubos de vacío, apareció sólo siete años más tarde, en 1958, y era diez veces más rápida: había comprimido sus componentes electrónicos en un espacio más pequeño que un naipe y era capaz de efectuar sumas en una millonésima de segundo. Los ordenadores de la tercera generación aparecieron a mediados de los sesenta. Contaban con circuitos integrados, eran aún diez veces más rápidos y en ellos los componentes electrónicos necesarios para sumar dos números en una millonésima de segundo cabían en un

espacio del tamaño de un sello de correos. Poco después aparecieron los chips, pequeñas plaquitas de silicio, con todos los componentes electrónicos de la batería de tubos de vacío embutidos en un pequeño círculo de metal. Pueden ser considerados «neuronas de silicio»(4). La cuarta generación apareció en los setenta y su *hardware* estaba constituido por circuitos integrados en muy larga escala. A principios de los ochenta se inició en Japón el programa de construcción de los ordenadores de quinta generación, que debían ser máquinas altamente inteligentes, capaces, en el futuro, de sustituir al hombre en la tarea de producir el conocimiento.

Toda esta tecnología se apoya en el denominado «sistema de símbolos», corriente largo tiempo dominante en IA. Según esta hipótesis cualquier sistema universal de símbolos con suficiente memoria puede llegar a ser, en virtud de su organización interna, adaptable a gran escala. Si se llega a la suficiente complejidad, la máquina llegará a ser capaz de interactuar con el mundo. Paralelamente, el llamado «modelo del dato» establece que el conocimiento puede ser representado por medio de símbolos, y este modelo sirve para todo conocimiento, de modo que cualquier conocimiento puede ser convertido en dato.

En su versión «fuerte», esta hipótesis afirma no sólo que el ordenador piensa, sino también que todo aquello que puede pensar es un computador. En consecuencia, la mente humana es un sistema universal de símbolos y todo pensamiento humano consiste en una manipulación de símbolos.

Los logros en esta dirección han sido deslumbrantes : el *Lógico teórico* (de Newell, Shaw y Simon) era capaz de demostrar treinta y ocho de los cincuenta y dos teoremas de los *Principia mathematica*.

Otro hito legendario fue *Eliza*, un programa diseñado con vistas a entablar una especie de diálogo en el contexto de una consulta psiquiátrica, en el transcurso del cual era capaz de emitir frases con sentido y, especialmente, solicitar más información del interlocutor. *Eliza* llegó a generar en seres humanos verdaderos lazos emocionales, hasta el punto de inquietar a su creador Joseph Weizenbaum.

Por su parte, el *General Problems Solver (GPS)*, llevó a sus creadores, Newell y Simon, a la conclusión, no ya de que las máquinas piensan, sino de que nosotros mismos somos computadores. En realidad, el *GPS* no pasa de ser el «mayor táctico de ensayo y error del mundo» (6). A partir de una posición inicial, recorre una lista de medios disponibles haciendo ensayos hasta llegar a una operación satisfactoria del propósito dado.

Un gran avance en el desarrollo de IA fue el que condujo desde el ordenador cuya efectividad se basa en su potencia de cálculo, hasta un sistema cuya eficacia reside en su capacidad de almacenar y utilizar conocimiento. Ello supuso una evolución desde el sistema algorítmico —que soluciona problemas concebidos como tránsito desde una situación inicial hasta una meta dada—, hasta la creación de sistemas en los que lo relevante es el conocimiento. Estos sistemas son los sistemas expertos (SE).

Los sistemas expertos (o sistemas basados en el conocimiento) fueron el campo más prometedor y brillante de todos los intentos ligados a la IA en los años ochenta.

Contienen dos elementos : una base de reglas y un motor de inferencia que selecciona las reglas apropiadas. De hecho, en muchas ocasiones han rivalizado con expertos humanos y con frecuencia los sustituyen. La aparición de los SE fue saludada como el inicio de una nueva época, dado que en la actualidad la riqueza se basa en el conocimiento y los ordenadores de «quinta generación» prometían ser realmente «inteligentes».

La estrategia de la ingeniería del conocimiento planteaba un conjunto de problemas relativos al conocimiento mismo. Si bien su almacenamiento no parecía plantear grandes problemas, no tardaron en aparecer otros, como el de cómo organizarlo y, muy especialmente, el problema del marco (o actualización de dichos conocimientos) y el problema de la pertinencia. La cuestión, pues, no estriba tanto en cómo acumular conocimiento sino cómo seleccionarlo en cada caso en función de su pertinencia. El escollo que esperaba a los eufóricos creadores de los ordenadores de quinta generación se encontró en la naturaleza misma de este conocimiento «experto» y en el manejo del conocimiento acumulado mediante el sistema de símbolos. Este conocimiento es, lógicamente, simbólico, explícito y descomponible en reglas. Ahora bien, no parece ser éste el sistema empleado por los seres humanos, por mucho que los defensores de la IA se empeñen en ello argumentando que si no lo entendemos así es simplemente porque nuestro procesamiento de la mayor parte de las reglas es inconsciente.

En realidad, como podremos ver más adelante, sólo desde una intención propagandística ha podido equipararse el sistema experto al experto humano. Hace ya dos décadas de la puesta en funcionamiento de DENDRAL y de MYCIN y los sistemas expertos están lejos de haber alcanzado sus expectativas. Los sistemas expertos son manuales automáticos de referencia sin comprensión de para qué sirven o cuáles son sus límites de aplicación (7) ; no comprenden ni pueden extrapolar soluciones a un contexto más amplio. Como dicen Benítez y Pérez de la Cruz, sólo son aplicables a dominios abstractos, formalizables, epistémicamente aislados. Éste es su «nicho ecológico», un micromundo de «eterno presente» en el que consiguen evitar cualquier referencia a una dimensión temporal (8). Como mucho, pueden advertir que si a es anterior a b y b anterior a c , entonces a es anterior a c . El sistema no sabe nada del significado de *anterior*, ni siquiera sabe que está ordenando los acontecimientos en el tiempo, y no en el espacio.

Mención aparte merece el proyecto CYC (Guha y Lenat, 1990 y 1993), un intento de romper los límites de aplicabilidad y conducir a los sistemas expertos al ámbito de la vida humana, que implica una portentosa acumulación de datos (está basado en el sistema de símbolos y en la hipótesis del dato), pero también de conocimiento sobre cómo utilizarlos, la elaboración de lógicas no monotónicas, la emulación del funcionamiento real de la mente y aun del sentido común humano. Se trata de uno de los proyectos más bellos y ambiciosos de la historia, de la primera incurción de la humanidad en ingeniería ontológica. Se trata de acumular los conocimientos sobre el mundo actual que se hallan al alcance de cualquier persona relativamente informada de finales del siglo XX (realidad consensuada) y de incluir además en el

programa un conjunto de reglas sobre cómo utilizar este conocimiento como trasfondo. La «ingeniería ontológica» supone la creación de lógicas no monotónicas, pero también la definición de qué son los objetos, y cuántas son sus clases. En suma, una memoria de conocimiento de sentido común en una base de datos que contiene todos los hechos de una enciclopedia, lo que exige fabricar millones de representaciones a partir de las cuales se espera que la máquina de inferencia sea capaz de deducir cualquier cosa a medida que se encuentre con novedades. Los problemas que el proyecto va encontrando a su paso no pueden calificarse de inesperados: el problema del marco (actualización de conocimientos), el de la pertinencia (organización que permita al programa seleccionar datos entre varios cientos de millones, entre los cuales se deberá establecer el grado de proximidad temporal, espacial e informacional, etc.), y el problema del filtro. Acaso la barrera de la complejidad sea insalvable, no de derecho para el sistema de símbolos, sino de hecho por la incapacidad humana para programar el conocimiento del mundo y el sentido común.

En los últimos años se está produciendo un cambio de paradigma propiciado por las desilusiones provocadas por la IA clásica, las computadoras digitales seriales del tipo Von Neumann. Se trata de la aparición de una alternativa radical, constituida por el conexionismo (procesamiento distribuido en paralelo, PDP), una innovadora apuesta por el desarrollo de redes neuronales (9) para la adquisición del conocimiento.

Esta opción parece suponer el abandono de la hipótesis del sistema de símbolos y el intento de construir *hardware* inspirados en el tejido neuronal humano. Si el cerebro es un mecanismo en paralelo más que en serie, en el que simultáneamente se producen millones de procesos neuronales, la simulación de las actividades humanas debería efectuarse mediante computaciones en paralelo. Se trata, pues, de reflejar la organización y el funcionamiento del cerebro humano, lo que ha motivado que los modelos conexionistas reciban el nombre de sistemas neuromórficos. La red neuronal intenta reflejar esa estructura de las salidas de las neuronas. Un conjunto de unidades de computación con diferentes valores de activación y con papeles de excitación e inhibición, masivamente conexionadas y operando en paralelo. Las redes neuronales operan en paralelo y su almacenamiento de información es distribuido, como parece ocurrir con las redes neurales cerebrales. Una máquina conexionista imita el cerebro humano o, más exactamente, las redes de neuronas, y además procesa la información en paralelo y no de modo serial o secuencial.2.

2. *Inteligencia artificial y concepción del hombre*

Si bien es cierto que a la ingeniería informática no tiene por qué preocuparle la relación entre tecnología y antropología, no lo es menos que, como en otras ocasiones, el desarrollo tecnocientífico está siendo acompañado de una redefinición de lo humano. Este paso desde lo tecnológico a lo ideológico, aunque con frecuencia acompañado de una renuncia explícita a todo lo que recuerde a filosofía, tiene consecuencias de importancia para la noción que sobre sí mismo tiene el hombre del

final del siglo XX, que se refieren a sus capacidades y posibilidades e incluso a la naturaleza y valor de lo humano como tal (10).

En su corta existencia, la informática ha llegado a impregnar todas las actividades del hombre y a reformular muchas de sus concepciones sobre las cosas y el mundo, penetrando una tras otra en todas las facetas del tejido social, en una evolución guiada por el «imperativo tecnológico» (11), *i.e.*, un desarrollo automotivado y, finalmente, descontrolado (12).

En la terminología de Shallis, se trata de una tecnología no protésica, sino sustitutiva del ser humano, al que se impone en virtud de su enorme rapidez y eficacia. Ante ésta e innumerables comprobaciones de la «superioridad» del ordenador, los hombres son progresivamente desalojados de puestos de responsabilidad —responsabilidad que es transferida a la máquina— o, en otros casos, se ven obligados a entablar con ella una relación simbiótica.

De su imponente poder transformador síguese su no menos formidable capacidad de fascinación. De hecho, no hay ni ha habido nunca una tecnología neutral. El despliegue de una tecnología exige la emergencia de unos valores e implica la eliminación de otros, como ocurrió con la difusión del reloj mecánico —para L. Mumford— o la imprenta —según Marshall McLuhan.

Como en el mito de Narciso, la tecnología informática ha llegado a erigirse en el modelo de la inteligencia y en personificación de la eficacia, la exactitud y el progreso mismo. Las máquinas han llegado a ser ante nosotros seres con los que compartimos lo que hasta aquí habían sido nuestros caracteres más «humanos» (14). Si en el plano tecnológico se da un *desplazamiento* del hombre en favor del ordenador, en el plano ideológico se llega a una *asimilación* de uno y otro. En efecto, si el cerebro humano es definible como una especie de máquina procesadora de información, entonces el hombre mismo puede ser entendido como una especie de máquina. Esta metáfora computacional es cada día más aceptada como descripción de lo humano y cuanto más espectaculares parecen los resultados del computador, más pequeños resultan ser los atributos de lo humano. En fin, la metáfora es interactiva, de modo que al decir que los ordenadores piensan afirmamos que los seres pensantes, los hombres, poseen los atributos de los ordenadores.

Distinguiremos tres aspectos de esta equiparación ideológica entre hombre y máquina: la concepción de la realidad y del conocimiento, el de la consideración del lugar del hombre en el proceso evolutivo y, finalmente, el de la concepción de lo que es el pensamiento. Vamos a considerarlos breve y sucesivamente.

En primer lugar, la equiparación entre el hombre y la máquina viene propiciada por la tendencia reduccionista con que nació la ciencia moderna, que, una vez dominante en el terreno de la *res extensa*, se ha creído llegada la hora de extenderla al de la *res cogitans*. Este reduccionismo constituye el norte mismo del avance de la ciencia, tiene su expresión más lograda en la física cuántica y de él es una muestra el archiconocido *Computing machinery and intelligence*, del que no tardaremos en ocuparnos.

El reduccionismo posee una doble vertiente, epistemológica y ontológica. En lo que se refiere al lado ontológico, la concepción reduccionista de la mente se fundamenta en la llamada «teoría de la identidad psiconeural» de Herbert Feigl, para quien lo psicológico se reduce a lo neuronal y lo neuronal a lo fisiológico. Las leyes del universo son físicas. Lo psíquico no constituye ninguna excepción para el monismo fiscalista (o «materialismo prometedor», al decir de Popper), pues sólo lo material es real.

Por otra parte —y hablamos ahora del lado epistemológico—, si todo es material y la materia es física, la Física será *la* ciencia a que deberán reducirse todas las demás. Esta especie de monismo epistemológico conlleva una aplicación discutible del «principio de economía» y sigue constituyendo una especie de supuesto filosófico, consciente o no, de muchos teóricos de la IA. Marvin Minsky nos proporciona un ejemplo llamativo, cuando se refiere con impaciencia («no soporto tales argumentos») (15) a ciertos filósofos para quienes la conciencia constituye una facultad básica que sólo puede poseer una mente viva. «A mi entender, —dice Minsky— tales asertos indican “envidia de la física”, celos de la perfección con que la física ha podido explicar tantas cosas a partir de un número muy reducido de principios» (16).

Desde una perspectiva evolutiva de la relación entre hombres y ordenadores, segundo aspecto del que prometíamos ocuparnos, se ha intentado establecer el lugar que corresponde a unos y a otros en el curso general de la evolución, haciendo de las máquinas nuestros «sucesores». N. Wiener (y también Von Neuman) ya había hablado de un futuro «que como mejor se describe es con la palabra pos-biológico (...) Es un mundo donde la raza humana ha sido barrida por la marea del cambio cultural, cuyo puesto ha sido usurpado por su propia progenie artificial» (17). Bruce Mazlish, un investigador del MIT, afirma que la aparición del ordenador representa la superación de lo que llama «la cuarta discontinuidad», la discontinuidad entre hombres y máquinas. La naturaleza es un continuo y la discontinuidad postulada por el antropomorfismo ha sido progresivamente desterrada de la ciencia. En este progresivo abandono de las pretensiones humanistas la revolución de los ordenadores constituye el cuarto *shock*.

En este orden de cosas, el bello libro de R. Jastrow, *El telar mágico* (18), atiende al lugar que corresponde al cerebro humano en el marco temporal de la evolución de la naturaleza y la vida. Para el prestigioso físico y astrónomo somos los últimos en esta sucesión, pero en modo alguno podemos afirmar que seamos el final de la historia. El hombre ha de tener su sucesor en «una inteligencia no biológica, brotando de la estirpe humana y destinada a superar a su creador» (p. XI). El hecho de que el cerebro, tras un millón de años de crecimiento explosivo, haya dejado de expandirse, sugiere que la historia de la evolución humana puede haber terminado y es razonable suponer que no seamos la última palabra en el desarrollo de la inteligencia sobre la Tierra. «Si el pasado sirve de alguna guía para el futuro, la humanidad está destinada a tener un sucesor más inteligente» (p. 145). «En una perspectiva más amplia, yendo más allá de la Tierra y sus criaturas biológicas, los auténticos atributos de la vida inteligente los encontraremos en aquellos que son compartidos por el

hombre y el ordenador» (p. 170).

De hecho, la aparición de la vida sobre la Tierra supuso una verdadera explosión de información. De la información genética a la extragenética (memoria en el cerebro) y de la extragenética a la extrasomática o cultural, el volumen y la calidad de la información de nuestra especie no ha hecho sino crecer. Si, además, reparamos en la escala de secuencias temporales de la evolución de la vida y la información, el panorama resulta muy expresivo: «el aumento de información posee un carácter explosivo : presenta avances que tardaron millones de años, miles de años, cientos de años y decenas de años en producirse (a través de lo que yo llamo cría artificial, los humanos han producido al menos cinco generaciones de ordenadores)» (19).

Pero el orden de consideraciones que resulta decisivo en la equiparación del hombre a la máquina es el que se refiere a la afirmación de que efectivamente los ordenadores piensan. La tesis en cuestión tuvo su origen en quien es sin duda uno de los hombres más importantes del siglo XX, el genial matemático inglés Alan Turing.

A. Turing había establecido que si es posible enunciar con precisión los pasos necesarios para realizar una tarea intelectual, de modo que queden formulados con nitidez sus elementos y las reglas para su solución, entonces esta tarea puede ser programada y ejecutada mecánicamente por un ordenador. Cualquier tarea de cómputo enunciada explícitamente era realizable por una máquina, siempre que poseyera un conjunto finito apropiado de instrucciones. El proceso computacional es capaz de procesar cualquier tipo de programa expresado en el código binario de unos y ceros, alcanzando la solución de cualquier problema de esta índole que se le plantee.

Por su parte, el «test» de Turing probaba que es posible sustituir al hombre por el ordenador en un juego en el que el interrogador es sistemáticamente engañado primero por el hombre y después por el ordenador. En la medida en que la suya fue una posición conductista (20), no definía lo que es pensamiento ni lo que es máquina, limitándose a proponer un juego. Para él y sus seguidores, la cuestión está bien planteada y resuelta si se admite que lo que hace la inteligencia es procesar información del mundo; este procesamiento puede realizarse con reglas meramente formales, expresables como algoritmos, por lo que, en realidad, lo que hace el cerebro es una especie de algoritmo. Ahora bien, cualquier algoritmo puede ser ejecutado en una máquina de Turing, por lo que la máquina es equiparable al cerebro. Por otro lado, negar que la máquina carece de conciencia es injustificado, ya que de hecho suponemos conciencia en nuestros interlocutores únicamente por analogía con los propios procesos mentales, ya que no podemos hacer otra cosa que suponer en ellos lo que ocurre en nosotros. En tal caso, ¿por qué negar que pueda ser autoconsciente la máquina? ¿por qué no suponer en ella un proceso similar al que se da en nosotros mismos? Tampoco impresionaba a Turing la objeción de que las máquinas no pueden crear, prerrogativa característicamente humana; para el matemático inglés, ningún cerebro es, en realidad, creador sino sólo combinador. En fin, Turing, atrincherado en una posición «conductista» y, por tanto, al abrigo de cualquier objeción que no tome como referencia lo puramente observable, afirma que la

máquina efectivamente piensa si nos lo hace creer. Vale decir, si nos engaña como si fuera un hombre, entonces es que su engaño vale lo que el de un hombre.

Nótese que el test implica dejar de lado muchas cosas que son relevantes tanto en el proceso de elaboración de un pensamiento como en su comunicación y recepción, pero ello no hizo pestañear a casi nadie dentro de la comunidad científica reunida en torno a IA : la máquina piensa puesto que nos puede engañar. Aparte de otras objeciones, cabe también preguntarse por la validez del test como tal. De hecho no lo supera un chimpancé, pero tampoco *Eliza* lo consigue. ¿Significa esto que la máquina no piensa, puesto que no es capaz de pasar la prueba? Parece, más bien, que lo que hace *Eliza* no es sino comparar patrones, atender mecánicamente a una sucesión de símbolos para ver si aparece un determinado patrón (palabra o grupo de palabras). El programador ha *asociado* a cada patrón un conjunto de oraciones y cuando el programa detecta un patrón en la entrada, escribe la primera oración de la lista que se ha asociado a ella. En caso de que no aparezca ninguno de los patrones almacenados, selecciona una oración de la lista de últimos recursos. Es injusto decir que piensa y lo es decir que no piensa.

Reparemos ahora en la necesidad de emplear terminales en la prueba en lugar de establecer una confrontación directa a lo largo de la comunicación, en el hecho de que las palabras tengan que aparecer en una pantalla, etc. Ello no constituye un detalle sin importancia, sino que denuncia la necesidad que tiene el test de descartar todo lo que recuerde a ser humano porque sólo así puede establecer una equivalencia entre funcionamiento de la máquina e interacción humana.

En fin, el concepto de pensamiento que se está manejando en todo ello sólo es aceptable si se redefine al ser humano en términos computacionales y se opta por desconocer radicalmente la vida, la individualidad, la corporalidad y la propia espiritualidad presentes en el hombre. Únicamente así *pensamiento* equivaldría a *inteligencia*. Sólo puede afirmarse que las máquinas piensan desde la reducción del pensamiento a algo tan carente de connotaciones humanas que ya no parece pensamiento, al habersele extirpado artificialmente conciencia personal, experiencia o conocimiento, esenciales en toda nuestra vida mental.

La crítica más interesante a la versión «fuerte» de la IA, que sostiene que una computadora es comparable a una mente humana, entiende y experimenta estados cognitivos, procede del escrito de J. Searle «Mentes, cerebros y programas» (1980), cuyo debate constituye tal vez una especie de síntesis de los principales temas de la IA (21). En el segundo capítulo de *Mentes, cerebros y ciencia* (22) define esta versión de la IA como defensora de la afirmación de que «el cerebro es solamente un computador digital y la mente es un programa de computador». «La mente es al cerebro lo que el programa es al *hardware* del computador»(p. 33), desatendiendo el aspecto biológico de la mente humana. Desde este punto de vista, cualquier sistema físico dotado de un programa con los *inputs* y *outputs* correctos tendría una mente en el mismo sentido que la tiene el hombre..

Searle recurre a un ejemplo provocativo : a un individuo que está encerrado en un cuarto se le entrega un fajo de papeles escritos en chino, idioma que desconoce,

hasta el punto de no distinguir sus caracteres de otros garabatos. Se le entrega después un segundo conjunto de caracteres chinos, junto con una serie de reglas para cotejar la segunda tanda con la primera, éstas últimas en un idioma que el sujeto puede interpretar. Las reglas le enseñan a poner en relación dos series de símbolos y el sujeto las aprende hasta correlacionar los caracteres entre sí, de modo que responde siempre con el conjunto correcto de caracteres si se le ha dado antes otro conjunto de caracteres. Más tarde se le formulan preguntas y respuestas en su idioma. El sujeto, con el tiempo, ha llegado a dominar las instrucciones para manipular los símbolos chinos hasta el punto de que desde fuera de la habitación las respuestas no se pueden distinguir de las que daría un hablante chino. Es más, las respuestas que da en su lengua y las que da en chino son igualmente correctas. Pero en el segundo caso se elaboran manejando símbolos formales sin interpretar. Pues bien, en lo que se refiere al chino, el sujeto se está comportando como la computadora cuando realiza operaciones computacionales sobre elementos formales específicos. Y es que los símbolos no tienen significado alguno, carecen de contenido semántico. Ahora bien, tener una mente es mucho más que tener procesos formales o sintácticos. «La mente tiene más que una sintaxis, tiene una semántica. La razón por la que un programa de computador no pueda jamás ser una mente es simplemente que un programa de computador es solamente sintáctico, y las mentes son más que sintácticas. Las mentes son semánticas, en el sentido de que tienen algo más que una estructura formal : tienen un contenido» (p.37).

La prueba de la máquina de Turing no puede establecer si un programa de computadora tiene mente o no. Cualquier individuo o máquina al que se suministre una serie de reglas formales para seguirlas en circunstancias específicas puede simular el funcionamiento de una mente. Mientras se trate de realizar operaciones sintácticas, no se puede afirmar que el sujeto comprende verdaderamente. Puesto que la computadora es una máquina de realizar operaciones formales, no necesita ningún conocimiento semántico, ni ningún saber acerca del mundo real, ni ninguna intención, para alcanzar ciertos efectos mediante su respuesta específica. Se trata, pues, de una entidad fundamentalmente distinta de un ser humano, el cual es capaz de comprender el contenido semántico de una emisión, persigue finalidades propias y se comunica.

Para Searle el factor decisivo es la intencionalidad, la propiedad de los estados y sucesos mentales por la cual éstos son dirigidos hacia objetos y circunstancias del mundo. Los estados intencionales incluyen creencias, temores, deseos e intenciones. Cuál sea la explicación científica y el sustrato biológico de este fenómeno, es algo que está por averiguar.

3. Inteligencia artificial y pensamiento humano

Importa ahora atender a la singularidad de lo humano, precisamente desde la perspectiva abierta por las reflexiones en torno a la IA. En este sentido, el trabajo de 1972, ya clásico, de M. Boden (23), destaca la diferencia entre mostrar las analogías

entre mentes y máquinas, y atribuir a los ordenadores las categorías de *subjetividad*, *significado* y *propósito*, tal como se entienden habitualmente, que sólo se pueden atribuir a los artefactos en un sentido secundario. (p. 546).

Joseph Weizenbaum, en *Computer Power and Human Reason* (1976) (24), niega con rotundidad que el hombre sea una máquina. Si, ciertamente, procesa información, este procesamiento no es equiparable al de un ordenador, que será siempre incapaz de simular la inteligencia humana. Weizenbaum subraya con acierto que hablar de inteligencia abstracta es justamente lo que ha originado el estéril debate acerca de si es posible construir ordenadores inteligentes. El concepto de inteligencia abstracta carece de sentido ; para llenarlo de contenido es necesario un punto de referencia, la especificación de un dominio de pensamiento y acción, gracias al cual adquiere significado.

Sentado esto, Weizenbaum señala cuatro limitaciones de los ordenadores frente al hombre : en primer lugar, la capacidad de obtener conocimiento a partir del lenguaje natural. En segundo lugar, el ser humano dispone de un conocimiento adquirido en virtud de su instalación en un cuerpo que ningún otro ser llegará a adquirir de la misma forma. En tercer lugar, hay cosas que el individuo humano aprende exclusivamente como resultado de un trato con seres humanos, por quienes ha sido tratado como tal. Aun suponiendo que un ordenador desarrollara autoconciencia y fuera capaz de modificarse en virtud de sus experiencias, la socialización humana seguiría siendo inevitablemente distinta de la de la máquina. En cuarto lugar, la información de un mensaje no es sólo una función del mensaje mismo, sino que depende del estado de conocimiento y de las expectativas del receptor. En el lenguaje humano se encuentran depositadas la historia de quienes lo utilizan, la de la sociedad y de la humanidad entera. Por el contrario, el lenguaje de los ordenadores es sólo un lenguaje funcional que identifica hechos y palabras únicamente con las metas inmediatas que han de lograrse o con los objetos que deben ser transformados. Estos cuatro no sólo son otras tantas limitaciones de los ordenadores, sino que constituyen rasgos diferenciales de la inteligencia humana frente a cualquier otra posible forma de inteligencia.

Más recientemente, R. Sokolowski, en «Inteligencia natural e inteligencia artificial» (25) contrapone la capacidad de razonamiento deductivo inferencial, que suponemos a la máquina, con otras formas de actividad mental más genuinamente humanas. Las deducciones lógicas pueden llegar a ser tan explícitas que se han vuelto susceptibles de ser realizadas mecánicamente, pero sería un error afirmar que ésta es la única clase de inteligencia. Hay otras dos formas de actividad intelectual no reductibles a la derivación, que son la capacidad de citar y la capacidad de distinguir. Por otro lado, Sokolowski subraya el papel de los deseos en la realización del pensamiento.

La capacidad que tenemos de citarnos unos a otros es irreductible a la lógica inferencial, anclada siempre en un único punto de vista. Esta restricción de la lógica es una abstracción útil y legítima, pero no proporciona una imagen completa del pensamiento humano. La lógica de la deducción es una lógica para monólogos, una

«lógica ciclópea». Toda diversidad de puntos de vista queda filtrada y en ella sólo se admite lo que se sigue de nuestras premisas. Incluso en las lógicas no monotónicas, que tratan de dar cuenta de situaciones y hechos que no se siguen de las premisas que se configuran en el sistema, seguimos limitados a inferencias que ejecutan desde un solo punto de vista. Por el contrario, nuestro pensamiento no deriva sus posturas sólo de axiomas aceptados como verdaderos porque cuenta con puntos de vista alternativos, que se niegan cuando nos concentramos en inferencias deductivas en línea recta.

Una segunda forma de actividad intelectual genuinamente humana es la de practicar «distinciones»: ciertamente, un programa de computadora puede hacer una distinción en el sentido de que puede seleccionar un ítem en lugar de otro, pero esa actividad presupone que los términos de la distinción ya han sido programados en la máquina. ¿Puede *originarse* una distinción en una máquina? En nuestro pensamiento natural no inferimos distinciones. Reconocer que hay dos aspectos distintos en una situación es un acto de pensamiento mas genuino que la inferencia.

En fin, tenemos la presencia de los deseos, que se hallan implicados con el pensamiento de dos maneras, el deseo de saber más y el deseo de otras satisfacciones como el alimento, el ejercicio, el reposo etc. ¿Cómo se relaciona con ellos el pensamiento? Una forma común de expresar esta relación consiste en decir que la razón es la esclava de las pasiones: los deseos proporcionan los fines, el pensamiento proporciona los medios.

Esa comprensión de la relación entre el deseo y la razón encaja bien con algunas presuposiciones de la IA pero implica que la computadora tendría que tener los objetivos configurados de antemano, así como necesitaría tener configurados sus axiomas. La computadora nos ayuda a obtener los fines elaborando las inferencias apropiadas al problema que enfrentamos y a los recursos que tenemos. De este modo, si la inteligencia natural es esclava de las pasiones, la inteligencia artificial puede emularla. Pero, a su vez, la razón natural no es por completo extraña a nuestros deseos. Deseamos de una manera pensante. Nuestras pasiones se hallan penetradas por la inteligencia y además surgen nuevas clases de deseos que sólo un ser pensante puede tener. Nuestro deseo racional genera no sólo curiosidad y la articulación pensante de las pasiones sino también el establecimiento de formas de desear que no se darían si no pensáramos.

La inteligencia artificial puede ser capaz de hacer algo con objetivos configurados de antemano, pero ¿puede emular la mezcla de deseo e inteligencia que constituye gran parte de lo que pensamos y hacemos? ¿Puede emular la curiosidad?

Si la inteligencia artificial fuera capaz de dar cuerpo a formas de pensamiento tales como la cita, la distinción y el deseo, parecería que de alguna forma es capaz de originar su propio pensamiento, de hacer algo que no puede resolverse en el razonamiento de quienes hacen y usan las máquinas pensantes.

Para Martínez Freire carece de sentido (y en ello coincide con Weizenbaum) una caracterización abstracta o absoluta de la inteligencia, y es preciso una estipulación concreta y relativa a las tareas que en el ser humano (punto de referencia histórico)

consideramos inteligentes. Sentado esto, para él resulta evidente que existe inteligencia artificial, ya que existen máquinas inteligentes, es decir, máquinas con programas para ejecutar tareas que en el ser humano atribuimos a su inteligencia. Pero no es menos cierto que las máquinas no consiguen igualar al ser humano en todas sus tareas inteligentes. El punto decisivo en esta cuestión es la (aparentemente) insuperable dificultad de las máquinas en igualar algo tan sencillo en el hombre como es su sentido común. En virtud del sentido común, el ser humano tiene unas capacidades intuitivas, que no parecen poder representarse en un procedimiento mecánico, de distinguir lo relevante de lo irrelevante en la información, de advertir lo que cambia y lo que permanece en una situación, y de revisar sus conocimientos y conclusiones según se producen nuevos datos. Precisamente buena parte de la investigación lógica actual en conexión con la inteligencia artificial dirige sus esfuerzos hacia los objetivos de formalizar y mecanizar el razonamiento del sentido común, habiendo surgido una veintena de nuevas lógicas.

Junto al sentido común, Martínez Freire subraya la existencia de otro rasgo diferencial decisivo entre la inteligencia artificial y la humana: la existencia de procesos «espirituales» en los seres humanos. Los procesos mentales de autoconciencia (o conciencia de si mismo como sujeto de los procesos mentales), los procesos de formación de un proyecto vital personal y las voliciones libres o indeterministas parecen ser características exclusivas de los seres humanos. «En resumen, podemos decir que algunas máquinas dotadas de programas adecuados en efecto piensan y están dotadas de mente, aunque sus procesos mentales, superiores en algunas tareas a los procesos humanos correspondientes, no agotan la enorme variedad y capacidades de tareas de los procesos mentales de los seres humanos. El hombre es y seguirá siendo el paradigma de la inteligencia» (26).

Hubert Dreyfus ha llevado a cabo desde los años setenta una crítica importante de la IA (*What computers can't do*) (27). Recibida como una incursión intrusista de un filósofo en temas científicos, planteaba la imposibilidad de que los ordenadores piensen. Para demostrarlo argumenta en dos direcciones : a) la discusión de la cuádruple asunción de supuestos filosóficos por parte de los defensores de la IA, y b) la determinación de lo específicamente humano.

El mérito de Dreyfus fue instalar el núcleo de sus reflexiones justo allí donde se dejaban como autoevidentes ciertas «asunciones» que, como tales, no recibían ningún tipo de justificación. En lo que se refiere a la asunción biológica, consiste en admitir que el cerebro procesa información en operaciones discretas mediante un equivalente biológico de los interruptores de encendido y apagado. La asunción psicológica acepta que la mente humana puede tratarse como un mecanismo que opera sobre unidades de información según reglas formales. En lo que toca a la asunción epistemológica, significa la aceptación de que el conocimiento en su totalidad puede ser formalizado. Por último, una asunción ontológica: toda la información acerca del mundo puede ser analizable en términos de elementos determinados independientes de la situación.

Las cuatro asunciones constituyen otras tantas hipótesis carentes de comprobación y completamente discutibles: ni es cierto que el cerebro funcione como un computador digital (asunción biológica), ni es equiparable el procesamiento de información mecánico al humano (asunción psicológica). Respecto a los presupuestos más directamente filosóficos, la tesis de que toda la conducta humana tiene que ser reducible a hechos atómicos y reglas, parece plenamente refutada por la experiencia científica misma: la formalización de micromundos no pasa de ser una solución *ad hoc* incapaz de resolver el problema y no es posible formalizar el sentido común.

Esto último se relaciona con el segundo grupo de sus argumentos, el énfasis puesto en lo específicamente humano: la IA en ningún caso se ocupa de ciertos aspectos que subyacen a toda conducta inteligente, como el papel del cuerpo en la organización y unificación del conocimiento, el papel de la situación y el de propósitos y necesidades.

Más recientemente, en un trabajo en colaboración con su hermano Stuart Dreyfus, «Fabricar una mente versus modelar el cerebro: la inteligencia artificial se divide de nuevo» (28), estudia el sentido e importancia de las redes neuronales en relación con la dificultad de emular el sentido común humano. Merece la pena que nos detengamos, siquiera sea brevemente, en esta argumentación.

Más atrás nos hemos referido a la diferencia entre la IA clásica y el conexionismo, que en su tiempo rivalizaron y cuya pugna se saldó con la victoria de la primera y el abandono del segundo. Si la primera buscaba utilizar las computadoras para instanciar (¿?) una representación del mundo, la segunda pretendía simular interacciones de neuronas. Para la primera, que se orientaba a la representación simbólica del mundo, mentes y computadoras son sistemas físicos de símbolos, en el sentido en que se expresaban Newell y Simon. Dicen en este sentido los Dreyfus que la hipótesis del sistema de símbolos llevaba las concepciones del primer Wittgenstein a un terreno empírico y basaban en ella un programa de investigación. Era, pues, una línea de investigación racionalista y reduccionista.

Ambos programas de investigación poseían ideas dignas de explorarse; ninguno de ellos podía considerarse indiscutible, pero indudablemente los éxitos que se estaban produciendo, como la prueba de teoremas lógicos o la resolución de enigmas combinatorios, jugaban a favor del sistema de símbolos. Sin embargo, para los Dreyfus, estos éxitos, por sí mismos, no explican el abandono radical del conexionismo; entraba en juego una concepción teórica en la que los defensores de la IA clásica se hallaban fuertemente atrincherados: la tradición racionalista y atomista contra la que poco pueden hacer (entonces y ahora) (29) quienes mantengan una concepción diferente del conocimiento (30). Esta tradición no sólo se remontaba al primer Wittgenstein, Descartes o Leibniz, sino a la visión racionalista del conocimiento propia de la tradición occidental.

Su estrategia transfería a todos los dominios los métodos desarrollados en las ciencias naturales: todos los dominios debían ser formalizables y la IA sólo tiene que encontrar elementos y principios *independientes del contexto*. Pero el problema —dicen los Dreyfus—, no tardó en aparecer; era, como dicen ellos, la «venganza del

mundo cotidiano», rebelde tanto ante la IA como lo había sido siempre ante la tradición racionalista. El problema que planteó a la IA atravesó tres etapas.

Entre 1955 y 1965 se ocupó de dos temas de investigación, la representación y la búsqueda, llamada entonces «simulación cognitiva». Newell y Simon demostraron cómo resuelve una computadora una clase de problemas ateniéndose al principio heurístico de búsqueda general conocido como análisis de medios y fines.

La segunda etapa (1965-1975), capitaneada por Marvin Minsky y Seymour Papert en el MIT, estaba relacionada con los hechos y reglas a representar. La idea consistía en desarrollar métodos para tratar sistemáticamente el conocimiento en dominios aislados llamados «micromundos». Se esperaba que estos micromundos, restringidos y aislados, fueran cada vez más realistas y que se combinaran de manera tal que llegaran a asimilarse al sentido común, pero los investigadores confundieron dos dominios que, siguiendo a Heidegger, distinguiríamos como «universo» y «mundo». Un conjunto de hechos interrelacionados puede constituir un universo, como el universo físico, pero no constituye un mundo. Este último es un cuerpo organizado de objetos, propósitos, habilidades y prácticas sobre cuya base adquieren sentido las actividades humanas.

En su tercera etapa, más o menos desde 1975 hasta el presente, la IA se ha enfrentado al «problema del conocimiento del sentido común». Los dos períodos iniciales (la simulación cognitiva y los micromundos) se caracterizaron por el intento de eludir el problema del conocimiento de sentido común y por ver hasta dónde se podía llegar haciendo intervenir tan poco conocimiento como fuera posible. A mediados de la década de 1970, sin embargo, se vio que ya no había forma de eludirlo. Este problema del sentido común ha bloqueado el progreso en la IA durante la última década (31), que aparece así como lo que I. Lakatos llamaba un programa de investigación degenerante (32). Formular una teoría del sentido común demostró ser más arduo de lo que se había pensado. La raíz del problema se halla, según nuestros autores, en la concepción racionalista del conocimiento, según la cual comprender un dominio consiste en poseer una teoría de ese dominio, que formula las relaciones entre los elementos objetivos, independientemente del contexto. Esta convicción arranca en el platonismo, para el que en dominios teóricos como las matemáticas y la ética, son aplicables reglas o teorías aprendidas en otra vida, fuera del mundo cotidiano. De acuerdo con Heidegger, la filosofía tradicional se define desde el comienzo por su interés por los hechos del mundo, mientras «pasa por encima» del mundo como tal, lo que significa que la filosofía ha ignorado sistemáticamente desde el comienzo el contexto cotidiano de la actividad humana.

Es una ironía de la historia intelectual que el devastador ataque de Wittgenstein a su propio *Tractatus*, sus *Investigaciones filosóficas*, se publicaran en 1953, justo cuando la IA se situaba en la tradición abstracta y atomista que él había fundamentado. Como Heidegger, destacaba la importancia de las prácticas cotidianas y sostenía que no se podía tener una teoría del mundo de la cotidianidad. De ese modo, cuestionaban la tradición misma en que se basaba el procesamiento de información simbólica.

Wittgenstein había buscado en vano los hechos atómicos y objetos básicos que su teoría requería, pero abandonó finalmente toda filosofía racionalista porque creía que el análisis de las situaciones cotidianas en términos de hechos y reglas sólo es significativo en algunos contextos y para algunos propósitos. Los elementos escogidos ya reflejan los objetivos y propósitos para los que fueron focalizados. Cuando procuramos encontrar elementos independientes del contexto y libres de propósitos, como debemos hacerlo si pretendemos encontrar los símbolos primitivos con que alimentar una computadora, estamos tratando en realidad de liberar aspectos de nuestra experiencia de la misma organización pragmática que hace posible usarlos inteligentemente al afrontar los problemas de la vida diaria.

La «venganza del mundo cotidiano» consiste, pues, en la rebeldía del contexto y la contumacia del sentido común, resistentes ante la objetivación reduccionista procedente de la tradición, prestigiada por los éxitos de la ciencia positiva y reinante en la IA clásica.

Aún sabemos demasiado poco sobre los contenidos y la estructura del conocimiento de sentido común. La dificultad y la necesidad de afrontar este problema, así como el esfuerzo de investigación al que se refiere no son, sin embargo, algo nuevo. La fenomenología de Husserl fue justamente uno de esos esfuerzos de investigación. Husserl, efectivamente, había hablado de la «pesada concreción» del noema y de su «tremenda complejidad», concluyendo con tristeza, a la edad de setenta y cinco años, que había sido un principiante perpetuo y que la fenomenología era una «tarea infinita». Había llegado a tropezar con la cuestión central de la IA : ¿puede formalizarse el sentido común?

4. *El raciovitalismo : la vida y el pensamiento como función vital*

Constituye un acierto de la filosofía raciovitalista la perseguida y conseguida atención a la peculiaridad de la vida que la caracteriza en todas sus dimensiones. Una de las consecuencias de esta lealtad al fenómeno vital es precisamente la índole de los rasgos con que en ella se perfilan las nociones de sujeto y objeto, hombre y mundo, pensamiento y realidad, que impiden tanto la *traducción* del mundo a repertorio de datos neutrales como la *reducción* del hombre a sujeto abstracto, intemporal y descontextualizado.

Para Ortega no hay modo de eludir el hecho de que el hombre mismo es el único *terminus a quo* de toda elaboración teórica. Por tanto, no hay dato que no suponga una interpretación y no hay interpretación que no proceda de un sujeto. Su filosofía parte, por tanto, del hombre. Ahora bien, para el filósofo español, propiamente hablando, también «hombre» es una abstracción ; lo que no es elaboración teórica es el yo concreto que cada uno es. Este yo es, ante todo, su vida o, mejor, su vivir, su estar viviendo aquí y ahora. Este dato es anterior al pensamiento y a toda otra realidad. Por otro lado, tampoco hay nunca un vivir abstracto. Vivir es siempre realidad última, experiencia concreta e irreductible. La vida es punto de partida porque no hay ninguna otra realidad que se le pueda anteponer y porque cualquier

actividad la supone. Lo demás (el dato, el axioma, la deducción, la teoría, la ciencia misma) es hipótesis inventada por el hombre para vivir.

En esta vida (33) concreta e irreductible que es aquí y ahora cada uno de nosotros, se hallan implicados yo y circunstancia, sujeto y mundo. En este punto, el raciovitalismo afirma con enérgica radicalidad una concepción relacional que exige romper a un tiempo con conceptos tradicionales y prejuicios propios del sentido común. Trátase de renunciar a todo centralismo del sujeto o del objeto. De esta concepción relacional se sigue una serie de determinaciones mutuamente implicadas, que repasamos en apretadísima síntesis.

La primera es la *mutua indigencia* : la vida es *coexistencia* y *coimplicación* (34). El yo vive en una circunstancia y esta coexistencia no puede entenderse como una dependencia unilateral sino como interdependencia, huyendo de toda autarquía del sujeto o el objeto, del yo o la circunstancia, del hombre y del mundo.

Por su parte, el yo se encuentra también en dependencia del mundo, de modo que cuando hablamos del yo, no hablamos de un «éste» a diferencia de «otro», ni de un este hombre a diferencia de las cosas. Cuando hablamos del yo, nos referimos a todo, hombre, cosas, situaciones, en plena ejecución (36). Yo no soy mi vida, sino yo con las cosas. El yo se capta ligado al no yo, aunque sean irreductibles; es una mismidad en alteridad, cuyo dinamismo es centrífugo y centrípeto. No se trata, pues de yuxtaposición de yo y mundo, sino de plena coexistencia y mutua obligación (38).

Por tanto, sujeto y objeto no se definen por su mutua negación sino por su implicación, por su radical pertenencia y mutua dependencia (39).

El segundo carácter decisivo de la realidad que llamamos «yo» o ser humano es la *ejecutividad*, una ejecutividad no consecutiva sino constitutiva. Se trata de una existencia que consiste ante todo en actividad, exigida y acompañada por la sustancial inadaptación humana al mundo, su consiguiente necesidad de adaptación y su permanente aspiración. La actividad del ser humano es previa al pensamiento, que puede ser considerado como una función a su servicio ; «*cogito quia vivo*» —dice Ortega. Las cosas, por su parte, son originariamente *pragmata* (40), asuntos, campos pragmáticos, instancias (41).

En tercer lugar, esta ejecutividad se da en *libertad*, Vivir es faena poética que se realiza en la inexorabilidad de lo dado, de la circunstancia. Circunstancia y vocación son materia y forma de la vida. La circunstancia no decide mi vida, pero sí la define. Para que se una a mí en una unidad necesita un ingrediente : la vocación. «Vida significa inexorable necesidad de realizar el proyecto de existencia que cada cual es (...) La vida es constitutivamente un drama, porque es la lucha frenética con las cosas y con nuestro propio carácter por conseguir ser de hecho lo que somos en proyecto» (42).

Así pues, el hombre es su quehacer. No es ya que tenga que elegir permanentemente, sino que elige *quién* es y *qué* es. Autodeterminación, pues, pero también ontodeterminación, porque en él la libertad metafísica es el origen de su libertad psicológica : el hombre es así, según la conocida expresión orteguiana, *causa sui* (43), un ser que tiene que ganarse la vida incluso metafísicamente (44).

En cuarto lugar, la vida humana, actividad y elección incesante en una circunstancia, posee un sentido *dialéctico*. El pensamiento mismo sigue un método o proceso dialéctico. En un primer momento la vida se acerca a una realidad tomando de ella una «vista», pero pronto, en virtud de la ley del contraste, esta perspectiva se asocia a su antítesis. Nos detenemos entonces en los distintos aspectos y tomamos una vista de él, y seguimos pensando o pasamos al aspecto contiguo sin abandonar los anteriores, sino más bien integrándolos (45).

: Por otra parte, el hombre es altruista *a natiuitate*. El otro es una realidad primaria para él. No se trata de un mero cuerpo sino de algo mucho más relevante, que Laín Entralgo ha denominado «carne», es decir, un ser corpóreo dotado de expresividad. El otro es para mí patente y latente, está dotado de interioridad que creo similar a la mía y no en vano mi relación con ella se llama convivencia (46). Por otro lado, al otro *le soy*. Mientras la naturaleza nos tranquiliza porque no tiene opinión, el otro produce alerta por su carácter recíprocante : cuando lo descubro, descubro un *alter ego* que me puede corresponder. Descubro que además de mi vida y de lo mío hay lo otro absolutamente inaccesible. Un puro no yo, no como el mundo, sino un no yo que tiene yo.

Así pues, además de mi vida y de lo mío, está «lo otro», lo «no yo», y *el otro*. Y sólo en el trato con él va surgiendo el mundo objetivo, las cosas que además de serme y serle, son. Sólo a través del tu conozco realmente el yo. La llamada primera persona es precisamente la última en aparecer.

Este carácter dialéctico de la libertad humana tiene, obviamente, un carácter *temporal*.

La dialéctica temporal humana se concentra y se concreta en una *perspectiva*. En efecto, la vida humana se da siempre desde un espacio y un tiempo. La perspectiva del hombre sobre las cosas, lejos de constituir un factor de distorsión, es precisamente su organización (48).

De todo ello se sigue una caracterización del pensamiento y del conocimiento humanos muy alejada de los planteamientos de IA y que parte de la idea central de que el pensamiento es, ante todo, una función vital. En efecto, para el raciovitalismo el pensamiento es una reacción a la vida, una función vital «como la digestiva o la circulación de la sangre» (49), que se hace real desde una vida en una circunstancia (50).

El hecho de que el pensamiento sea una función de la vida nos lleva a la implicación de yo y circunstancia, una implicación que nos hace ver los conceptos y las cosas como íntimamente relacionados con el yo. Las cosas, pues, no tienen un ser en sí mismas : desarticuladas, en su ser para sí, son meras abstracciones, ficciones que supongo cuando finjo que no me ocupo de ellas. En su ser para mí, presentes en el plexo de mi circunstancia, y por tanto, en su ser originario, consisten en un «plan de atenuamiento». Su ser es interpretación, una interpretación hecha desde la ejecutividad de un ser que tiene que realizarse y que para ello necesita saber a qué atenerse «para intentar extraerse a sí mismo del enigmático pozo en que cayó al existir» (52).

Si el ser de las cosas es su interpretación, la vida es el intérprete. El hombre es una suerte de demiurgo que hace que el caos se convierta en cosmos y la realidad en sistema (53). La inteligencia distribuye las significaciones desde una *perspectiva*, en torno a un foco, creando un sistema en el que no caben cosas individuales. El principio unificador de este sistema no es otro que la preocupación humana (54) y el sistema mismo al que llamamos mundo es una respuesta creada por el hombre para afrontar su inseguridad.

Esta vida concreta anclada en el espacio y en el tiempo, dotada de una perspectiva desde la que construye su mundo o circunstancia, enfrentada a ella y en coexistencia con ella, condensa el sentido de sus experiencias en la forma de lo que llamamos sentido común. Ha sido necesaria toda la parvular inocencia de los ingenieros informáticos metidos a filósofos para pensar que el sentido común debería ser fácil de codificar e implementar en la máquina. El sentido común, el conocimiento general del mundo, supone un inmenso depósito de conocimientos ganados no de forma desinteresada, abstracta o teórica, sino a fuerza de experiencias orientadas siempre por una preocupación vital; lo presente y lo latente, la realidad del universo, fondo sobre el que se recortan de modo artificial las cosas, la conciencia de la perspectiva, la instancia de las cosas como campos pragmáticos, todo ello proviene de una realidad radical que es la vida humana, único punto de partida real y única capaz de hacer real un pensamiento humano

NOTAS

- (1) Sobre IA en general, J. Copeland, *Inteligencia artificial*, tr. J.C. Armero. Madrid, Alianza Editorial, 1996. E. Rich, y K. Knight, K. *Inteligencia artificial*, tr. P.A. González y F. Trescastro. Madrid, Mc Graw Hill, 1997.
- (2) Suelen distinguirse IA «ajena» y «humana», «débil» y «fuerte», «clásica» y «conexionista».
- (3) Una historia de los primeros treinta años de la IA, cf H. Gardner, «Inteligencia artificial: la herramienta del experto», en *La nueva ciencia de la mente*. Tr. L. Wolfson. Paidós, Barcelona, 1988, pp.159-203. Una visión actualizada se halla en Stuart Russell y Peter Norvig: *A Intelligence artificial modern approach*, Prentice Hall, Upper Saddle River, New Jersey, 1995
- (4) R. Jastrow, *El telar mágico*. Tr. D. Santos. Salvat, Barcelona, 1988, p.155
- (5) P. Martínez Freire, «El futuro de las máquinas pensantes», *Diálogo filosófico*, 35 (1996), pp. 234-250. También *La nueva filosofía de la mente*. Gedisa, Barcelona, 1995, pp. 103-4
- (6) J. Copeland: *Op. cit.*, p. 137.
- (7) «El término de sistema experto ha sido creado con fines de propaganda» (J. Weizenbaum, «La inteligencia artificial es considerada como una tendencia ideológica», *Antropos*.164 (1995), p.87. También. Hilary Putnam: «Mucho ruido por muy poco», en S. R. Graubard (comp.), *El nuevo debate sobre la inteligencia artificial. Sistema simbólicos y redes neuronales*. Trad. C. Reynoso. Gedisa, Barcelona, 1993. p. 308.
- (8) J. Copeland, *Loc. cit.*, p. 148
- (9) Cf Jack D. Cowan y D. E. Sharp: «Redes neuronales e inteligencia artificial», J. Schwartz: «El nuevo conexionismo: desarrollando relaciones entre la neurociencia y la inteligencia artificial» y G.B. Reeke (h) y G. M. Edelman: «Cerebros reales e inteligencia artificial», en S.R. Graubard, *Op. cit.*
- (10) Esa es la opinión de Herbert Simon. Cf J. Copeland, *Op. cit.*, p. 216. De modo más general, afirma D.C. Dennett: «La IA (...) a menudo está directamente involucrada en preguntas filosóficas inmediatamente reconocibles: ¿Qué es la mente? ¿Qué es el significado? ¿Qué son el razonamiento y la racionalidad?»

dad?...» «Cuando los filósofos se encuentran con la inteligencia artificial», en St. R. Graubard : *Op. cit.*, p. 320.

(11) Michael Shallis, *El ídolo de silicio*. Tr. J. Vicuña. Salvat, Barcelona, 1986, p. 86.

(12) J. Weizenbaum alerta sobre los riesgos de la improvisación de la carrera tecnológica y comercial, cuyo control se ha perdido incluso en los sectores comercial y militar. «Sencillamente no comprendemos nuestras máquinas» (p. 85), lo que puede empujarnos a vivir una catástrofe como la nuclear (J. Weizenbaum, *loc. cit.*, p. 85).

(13) Marvin Minsky, «¿Serán los robots quienes heredarán la Tierra?» *Investigación y ciencia*, diciembre de 1994 p. 90. Pascual Martínez Freire, *loc. cit.*, p. 247).

(14) Así, dicen Newell y Simon: «la intuición, sabiduría y aprendizaje ya no son propiedad exclusiva de los seres humanos : también se puede programar para que muestre estos rasgos cualquier computador de alta velocidad (*Apud* J. Copeland. *Loc. cit.*, p. 150).

(15) *Loc. cit.*, p. 90.

(16) Minsky llega a veces a la provocación. Weizenbaum dice de él : «...esta élite está formada en Estados Unidos únicamente por hombres. A veces pienso que el punto de partida de todos sus esfuerzos está en cierta envidia uterina». (*Loc. cit.*, p. 87).

(17) *Apud* B. Mazlish, *La cuarta discontinuidad*. Tr. : M. Arnáiz y A.L. Sanz. Alianza Editorial, Madrid, 1995, p. 321. Hans Moravec afirma que «nosotros, los humanos, los beneficiaremos de su trabajo durante algún tiempo» (...) pero tarde o temprano, como los niños naturales, se buscarán la vida mientras nosotros, sus ancianos padres, desapareceremos en silencio» (*Mind Children. The future of robot and Human Intelligence*. Cambridge. Harvard University Press). Marvin Minsky, en su artículo antes citado, se refiere a una «selección innatural» y a nuestra capacidad de diseñar nuestros «niños mentales». (*Loc. cit.*, p. 90).

(18) R. Jastrow, *El telar mágico*. Tr. D. Santos. Salvat, Barcelona, 1988.

(19) B. Mazlish, *Ibidem*.

(20) «Computing machinery and intelligence», *Mind*, (59) 1950 ; 434-60.

(21) La crítica no se dirige contra la versión débil de IA, que sostiene que la inteligencia artificial puede aclarar los procesos mentales y considera a la computadora una herramienta para el estudio de la mente.

(22) Trad. L. Valdés. Cátedra, Madrid, 1994.

(23) *Inteligencia artificial y hombre natural*. Trad. J.C. Armero. Tecnos, Madrid, 1984.

(24) Penguin Books, Londres, 1976.

(25) En St. R. Graubard (comp.) : *Op. cit.*, pp. 59-80.

(26) *La nueva filosofía de la mente*, p. 122.

(27) Manejamos una edición posterior, *What computers still can't do?*, de 1992. En 1986 él y su hermano Stuart publicaron *Mind over machine*. (Cf Martínez Freire, *La nueva filosofía de la mente*, p. 105 y ss).

(28) En St. R. Graubard (comp.), *Op. cit.*, pp. 25- 58.

(29) Gardner sugiere que el enfoque fenomenológico de la comprensión realizada desde un cuerpo corresponde a una «tradición epistemológica que es extraña virtualmente a todos los habitantes del mundo de la ciencia de la computadora y de la inteligencia artificial». (*Loc. cit.*, p. 184).

(30) Como veremos más adelante, ver las cosas de esta forma es ignorar el papel preeminente de la discriminación de patrones en la experiencia humana y el trasfondo de comprensión de sentido común que está presupuesto en el pensamiento cotidiano.

(31) T. Winograd ha sido uno de los primeros en ver las limitaciones de SHRDLU y de las dificultades para extender la estrategia de los micromundos. Perdida la fe en la IA, enseña ahora Heidegger en su curso de computación y señala «la dificultad de formalizar el fondo de sentido común que determina qué argumentos, metas y estrategias son relevantes y cómo interactúan».

(32) Cf Hubert y Stuart Dreyfus : *Mind over machine*, Oxford, Basil Blackwell, 1986.

(33) Cf J. Ortega y Gasset : *El hombre y la gente*. En *Obras completas*. Alianza Ed., Madrid, 1983, vol VII, p. 101 ; *La idea de principio en Leibniz*, vol. VIII, p. 274-5. ; *Qué es filosofía*, vol. VII, p. 411 :

(34) *La idea de principio en Leibniz*, vol. VIII 268.

(35) *Meditaciones del Quijote*, vol. I, p. 322 ; *Comentario al Banquete de Platón*, vol. IX, p. 768 ; *Prólogo para alemanes*, vol. VIII, p. 54 ; *Qué es filosofía*, vol. VII p. 431) ; en *El Quijote en la escuela* lo llama «paisaje de la especie» (vol. II, p.303), repertorio de posibilidades vitales (*Temas de viaje*, vol. II, p. 372). El mundo es «lo que nos afecta» (*Qué es filosofía*, vol. VII 416-7), lo vivido como tal (*Op. cit.* P. 424-5), «lo que nos es» (*Comentario al Banquete de Platón*, vol. IX, p. 768).

- (36) «Todo», mirado desde «mí» es el yo (*Ensayo de Estética a manera de prólogo*, vol. VI, p. 252).
- (37) J. Bayón, :*Razón vital y dialéctica en Ortega*, Revista de Occidente, Madrid, 1972, p. 22.
- (38) *La idea de principio en Leibniz*, vol. VIII, p. 268.
- (39) Cf *Apuntes sobre el pensamiento. Su teurgia y su demiurgia*, vol.V, p. 545), Cf R.Rodríguez : «Ortega y la reforma de la noción de ser». *Anales del Seminario de metafísica* (1983) pp. 35-49.
- (40) *El hombre y la gente*, vol. VII, p. 110. Cf J. Bayón, *loc. cit.* p. 32.
- (41) Cf A. Rodríguez Huéscar : *La innovación metafísica de Ortega*. Ministerio de Educación y Ciencia, Madrid, 1982.
- (42) *Goethe desde dentro*, vol. IV, p. 400.
- (43) *Historia como sistema*, vol. VI, p. 33.
- (44) Cf *El hombre y la gente*, vol. VII, 102 y ss.
- (45) *Origen y epílogo de la filosofía*, vol. IX, p. 374-375.
- (46) P. Laín Entralgo, *Teoría y realidad del otro*, Alianza Ed., Madrid, 1983, p. 235, 6.
- (47) El hombre es un ser *in via* (no factum sino faciendum0.) y la suya una temporalidad que expulsa la segmentación en instantes, ostentando un carácter procesual. Cf P. Cerezo Galán : *La voluntad de aventura*. Ariel, Barcelona, 1984.
- (48) Cf A. Rodríguez Huéscar : *Perspectiva y verdad*. Revista de Occidente, Madrid, 1966.
- (49) *El tema de nuestro tiempo*, vol. III, p. 164; *Prólogo a la «Historia de la Filosofía» de Karl Vorländer*, vol. VI 293. Cf F. Alluntis : «La razón vital en Ortega y Gasset». *Pensamiento*. (1983), pp. 421-435.
- (50) *Qué es filosofía?*, vol. VII, p. 423.
- (51) *La idea de principio en Leibniz*, vol. VIII, p. 277-8.
- (52) *Apuntes sobre el pensamiento. Su teurgia y su demiurgia*, vol. V, p. 529-30. La vida, para Ortega, contiene siempre un ingrediente de desorientación (Cf *Unas lecciones de Metafísica*, vol. XII. P. 49).
- (53) *Adán en el paraíso*, vol. I, p. 480.
- (54) *Meditaciones del Quijote*, vol. I, p. 312.

* * *

J.M. Atencia López
 Manuel Palacio, 5
 29017 Málaga