Ágora

El futuro de las máquinas pensantes

Pascual F. Martínez-Freire

¿Llegarán los ordenadores a sustituir a los seres humanos? Lo que hasta hace unas décadas parecía una pregunta de ciencia ficción, es ahora defendido con seriedad por algunos especialistas en Inteligencia Artificial. El presente artículo dibuja las más importantes líneas de investigación en este campo y las posiciones más relevantes respecto de la relación hombre-máquina pensante. La conclusión que propone el autor supone una decidida apuesta por un humanismo en el que quepa integrar el desarrollo tecnológico sin perder de vista sus innegables implicaciones éticas.

1. El mundo de los ordenadores

Para todos nosotros, o al menos para la mayoría, los ordenadores son ya algo familiar. Los ordenadores se han instalado en los bancos, en las oficinas públicas y privadas, en los comercios, en los despachos universitarios e incluso en nuestras casas bajo la modalidad de ordenador personal. En suma, los ordenadores forman ya parte integrante de nuestra vida cotidiana.

Por otro lado, sabemos o tenemos cierta información acerca del hecho de que los ordenadores son cada vez más «listos», hacen cada vez cosas más difíciles o más complicadas, son más flexibles en sus tareas, etc. Luego me ocuparé de analizar los distintos campos de desarrollo y las nuevas líneas de los ordenadores o computadores, pero de momento quiero destacar que, además de familiares, los ordenadores se presentan como algo en progresivo desarrollo.

Respecto de este progresivo desarrollo podemos detectar dos actitudes generales. En primer lugar, una actitud escéptica o pesimista, consistente en

sostener que por mucho que avancen los ordenadores nunca alcanzarán los niveles de racionalidad del ser humano y que, por supuesto, no conseguirán tener algunas características específicamente humanas como la creatividad o la sociabilidad y, mucho menos, la conciencia, emociones o el sentido moral.

En segundo lugar, podemos registrar una actitud optimista, consistente en creer que los ordenadores ya han alcanzado un nivel de desarrollo intelectual comparable al del ser humano y que, con el tiempo (y puede ser cosa de diez o veinte años), los ordenadores conseguirán ser creativos y constituir entre ellos una auténtica red de relaciones sociales. En cuanto a la conciencia, emociones y sentido moral, o bien se desvanecerán como falsos problemas filosóficos, o bien serán redefinidos (en términos mecanicistas) a la luz del desarrollo de los ordenadores.

Entre aquellos que aceptan este optimismo respecto del avance de los ordenadores, son posibles, a su vez, dos posturas distintas. Por una parte, algunas personas creen que esta competencia de los ordenadores puede ser peligrosa para la humanidad, que los ordenadores pueden llegar a volverse contra sus creadores y que, por tanto, debe interrumpirse la «carrera de los ordenadores» (tal como creo oportuno denominarla), de modo análogo a como debe interrumpirse la «carrera de armamentos», puesto que en ambos casos el desarrollo tecnológico resulta ser una amenaza contra la integridad o la supervivencia de los seres humanos.

Por otra parte, otras personas sostienen que la «carrera de los ordenadores» debe ser llevada hasta sus últimas consecuencias, que frenarla o interrumpirla supone frenar o interrumpir la propia evolución cultural y que, incluso si los ordenadores llegasen a sustituirnos como señores del mundo, debemos aceptar nuestra nueva situación de «especie» subordinada, contentándonos con el orgullo de haber creado a los nuevos señores de la tierra.

2. El nacimiento de la inteligencia artificial

En un principio, los ordenadores eran meros procesadores de datos, de gran tamaño, con escasa velocidad de cálculo y diseñados para ejecutar alguna tarea definida y concreta. Por ejemplo, el ENIAC (Electronic Numerical Integrator and Calculator), construido en 1946 en la Escuela Moore de Ingeniería Eléctrica de la Universidad de Pensilvania, que suele ser considerado como el primer computador, ocupaba toda una habitación y, cuando hoy vemos una fotografía suya, más nos parece una estación telefónica que un ordenador.

Pero la situación cambió rápidamente de los años cuarenta a los cincuenta y sesenta. Para empezar, los ordenadores disminuyeron progresivamente su tamaño, gracias al empleo, primero, de transistores y, luego, de circuitos integrados. Además, aumentaron su velocidad de cálculo así como su memoria, y adquirieron versatilidad, es decir, capacidad para ejecutar ta-

reas diversas, surgiendo con ello los ordenadores de propósito general (y no para un único propósito).

Más llevados por esperanzas (y el deseo de hacer ruido publicitario) que por auténticas realidades, en el verano de 1956 un grupo de diez matemáticos y lógicos americanos se reunieron en el Dartmouth College, en Hanover (New Hampshire). Su objetivo y denominador común era el desarrollo de máquinas inteligentes, es decir, el diseño de programas de ordenador que ejecutasen tareas que, hechas por el ser humano, calificamos de inteligentes.

En esta reunión estaban presentes los más destacados científicos que, por entonces, eran conocidos como contribuyentes en el campo del desarrollo de ordenadores con programas inteligentes. Arthur Samuel, que había diseñado programas para que un ordenador jugase a las damas, era uno de los asistentes. También estaban Allen Newell y Herbert Simon, quienes presentaron un programa para demostrar teoremas lógicos mediante ordenador. Otro participante era Marvin Minsky (del cual volveremos a hablar más adelante), quien ha sido y es desde entonces el más optimista y entusiasta propagandista de las excelencias y posibilidades de las máquinas pensantes o inteligentes. Finalmente, también debemos destacar la presencia de John McCarthy, quien precisamente acuñó la expresión «inteligencia artificial» para referirse a los trabajos que todos ellos estaban realizando.

Desde entonces, se entiende por inteligencia artificial la ciencia que estudia el diseño y construcción de máquinas inteligentes o pensantes, esto es, mecanismos capaces de realizar tareas que en los seres humanos atribuimos a su inteligencia, tales como demostración de teoremas, diagnóstico de enfermedades, análisis (sintáctico o semántico) del lenguaje, ganar una partida de ajedrez, etc.

Creo conveniente una aclaración terminológica. Acabo de emplear como sinónimos pensamiento e inteligencia, ya que he hablado acerca de máquinas inteligentes o pensantes. Algunos filósofos, siguiendo a Platón, podrán distinguir entre inteligencia como racionalidad discursiva (que se manifiesta en la creación de argumentos) y pensamiento como intuición racional (que se manifiesta en la comprensión o perspicacia). Aunque tal distinción puede ser legítima, e incluso útil, sin embargo no la adoptaremos, puesto que entiendo que tanto el discurso como la intuición racionales son dos aspectos de la actividad racional en general. Por ello emplearé como sinónimos máquina inteligente y máquina pensante, entendiendo además que tales máquinas desarrollan una actividad racional.

Otra aclaración, esta vez de carácter filosófico, es que no entraré ahora en la discusión de si las máquinas realmente piensan o simplemente imitan el pensamiento. De tal cuestión me he ocupado en mi libro *La Nueva Filosofía de la Mente* (Gedisa, Barcelona, 1995). Entenderé, a efectos prácticos, que una máquina piensa si se comporta como si pensase, es decir, adoptaré el punto de vista conductista propio del test de Alan Turing: si la tarea intelectual ejecutada por una máquina no puedo distinguirla de la tarea intelectual cumplida por un ser humano, entonces diré que esa maquina piensa.

Esta actitud conductista no debe parecer simplista, ya que habitualmente la aplicamos también a los seres humanos, pues, en efecto, creemos que las otras personas piensan porque se comportan como cuando nosotros pensamos, sin tener seguridad de que realmente piensen, salvo realizando análisis ulteriores acerca del funcionamiento de su cerebro.

3. El desarrollo de la inteligencia artificial

Desde los años cincuenta y, en particular, desde los años sesenta, el campo de la inteligencia artificial experimentó un notable desarrollo en distintas áreas de investigación y creación técnica. Enumeraré brevemente estos logros clásicos que agruparé en cinco apartados.

En primer lugar, se han creado programas de juegos para ordenador, tales como el juego de damas, tres en raya o el ajedrez. Ya hemos mencionado los programas del juego de damas de Samuel. En cuanto al ajedrez, deporte intelectual por excelencia, algunos campeones internacionales han sido derrotados en ocasiones por computadores con programas adecuados, tal como el Chess Champion Mark V. Actualmente, los ordenadores, al menos como jugadores de ajedrez, son serios rivales de los seres humanos, ya que máquinas como CRAZY BLITZ o bien HITECH ganan a cualquiera que no sea un maestro mundial.

En segundo lugar, se han desarrollado programas para demostrar teoremas lógicos y teoremas matemáticos. Así, el Logic Theorist, diseñado por Newell, Shaw y Simon (y publicado ya en 1956), prueba la mayoría de los teoremas contenidos en los *Principia Mathematica,* la obra lógica escrita entre 1910 y 1913 por A. Whitehead y B. Russell y que constituye la fuente básica de la lógica matemática clásica. Asimismo, por ejemplo, en 1976, se demostró mediante computador el famoso teorema de los cuatro colores; tal teorema, que durante más de un siglo los humanos intentaron demostrar sin éxito, establece que es posible colorear un mapa político usando sólo cuatro colores y de modo que los paises limítrofes tengan distintos colores.

En tercer lugar, también se han diseñado programas de ordenador para tratar la sintaxis y la semántica de los lenguajes naturales. Existen programas de traducción automática, como el programa creado ya en 1955 por Anthony Oettinger para traducir del ruso al inglés. Y también hay programas para analizar las frases del lenguaje natural en sus distintos componentes, o para buscar equivalentes semánticos. Sin embargo, las limitaciones de las máquinas en este área de investigación son importantes. En cuanto a la traducción, existe el serio problema de que el campo de ambigüedad de las palabras no coincide de un idioma a otro; por ejemplo, la palabra inglesa *date* que significa «cita entre amigos», también significa «dátil», cosa que no ocurre con la palabra española «cita». En cuanto a los análisis sintácticos, las reglas sintácticas de los idiomas poseen excepciones difíciles de recoger en un programa de ordenador, aunque veremos que recientemente se han producido nota-

bles avances. Y, en cuanto a los análisis semánticos, la ya citada ambigüedad, además de la dependencia del lenguaje respecto del contexto o situación de los hablantes, pone las cosas difíciles a las máquinas pensantes.

En cuarto lugar, otra área de investigación clásica en inteligencia artificial es la de reconocimiento de imágenes o visión computacional. En esta parcela se trata de crear programas que reconozcan o identifiquen rostros, huellas dactilares y, en general, objetos determinados. Ya en 1965, Roberts creó un programa para reconocer objetos tridimensionales que era capaz, por ejemplo, de identificar los objetos en una fotografía. La visión computacional es un área de gran importancia para la robótica, de la que hablaremos más adelante, puesto que para disponer de robots que puedan moverse con cierta libertad deben estar dotados de una capacidad de visión de la situación y de los objetos que manejan. Pero también aquí las máquinas pensantes encuentran limitaciones, ya que habitualmente los programas permiten al computador reconocer ciertos patrones de objetos bien definidos pero no otros patrones ajenos al programa ni tampoco objetos de escasa definición. Con todo, como veremos luego, actualmente existen nuevos programas que podrían vencer estas limitaciones.

Finalmente, en esta enumeración de las principales áreas de investigación en inteligencia artificial, me detendré en lo que suele considerarse la joya de tales investigaciones: los llamados «sistemas expertos». Un sistema experto, también denominado «sistema basado en conocimiento», es un programa de ordenador que permite a una máquina dar, para problemas específicos, soluciones sensiblemente iguales a las que daría el experto humano en tales problemas, pudiendo además la máquina justificar las soluciones ofrecidas.

Éste ha sido el campo tradicional de mayor éxito de la inteligencia artificial, ya que en él se iguala el comportamiento de una máquina a la capacidad de pensamiento de diferentes expertos humanos. Ya en 1964, Joshua Lederberg creó el sistema experto DENDRAL para realizar análisis químicos, sistema que fue perfeccionado en la Universidad de Stanford (California). Asimismo, en 1968, se diseñó el sistema MACSYMA, que llegó a emular a los expertos humanos matemáticos al poder realizar más de seiscientos tipos diferentes de operaciones matemáticas, incluyendo la diferenciación y la integración. Posteriormente aparecieron otros sistemas expertos clásicos e importantes. Por ejemplo, en 1972, el sistema MYCIN, empleado para el diagnóstico y tratamiento de las enfermedades infecciosas, o bien, en 1973, el sistema INTERNIST, que es utilizado para el diagnóstico en medicina interna. También cabe citar el sistema experto PROSPECTOR, que es usado para la exploración geológica. En suma, las máquinas dotadas de sistemas expertos pueden sustituir a expertos humanos tan diversos como matemáticos, geólogos, médicos o químicos.

En un sistema experto hay dos elementos básicos: una base de reglas y un motor de inferencia. Las reglas de la base son generalmente del tipo, por poner un ejemplo sencillo, «si tiene más de 37 grados, entonces tiene fie-

bre». La base de reglas intenta reunir un conjunto de reglas que codifiquen el conocimiento y la experiencia de un experto humano en un área concreta. A su vez, el motor de inferencia selecciona las reglas apropiadas para una situación dada y, mediante su manejo, llega a una conclusión. Además, el sistema puede justificar tal dictamen o conclusión, explicitando el conjunto de reglas de la base que ha usado.

4. El estado actual de la inteligencia artificial

Actualmente los científicos que trabajan en inteligencia artificial suelen adoptar una de las siguientes posturas. Para algunos, las cosas van bien y con las técnicas clásicas podemos progresar indefinidamente. Para otros, los avances previstos en los años ochenta no se han producido ni se producirán y la inteligencia artificial ha quedado estancada. Finalmente, para un tercer grupo, la inteligencia artificial debe cambiar radicalmente sus métodos y técnicas y, merced a ello, nuevos progresos de gran alcance son posibles.

Los científicos conservadores y satisfechos en inteligencia artificial hacen suya la distinción trazada por el filósofo John Haugeland, en *Artificial Intelligence: The Very Idea* (1985), entre GOFAI y otras posibles alternativas de investigación en inteligencia artificial. En efecto, GOFAI es un acrónimo para «Good Old Fashioned Artificial Intelligence», que podemos traducir como «la buena inteligencia artificial a la antigua usanza». Para estos científicos conservadores y satisfechos, la inteligencia artificial clásica, iniciada en los años cincuenta y desarrollada a partir de los sesenta, constituye un buen paradigma de investigación y no son necesarios paradigmas alternativos.

Para los científicos pesimistas en inteligencia artificial constituye una referencia ejemplar el anuncio y posterior abandono de la llamada quinta generación de ordenadores por parte de las autoridades científicas japonesas. Podemos distinguir cuatro generaciones básicas de ordenadores: la primera generación con su *hardware* (esto es, la constitución física de las máquinas) de tubos de vacío, como el célebre ENIAC, la segunda generación con *hardware* de transistores, la tercera con *hardware* de circuitos integrados y, finalmente, la cuarta generación con circuitos integrados en muy larga escala.

Pues bien, a principios de los ochenta, las autoridades científicas japonesas anunciaron su decisión de financiar y trabajar en una «quinta generación», que estaría constituida por máquinas altamente inteligentes. Tal anuncio fue recibido con entusiasmo por muchos científicos y filósofos. Así, por ejemplo, Edward Feigenbaum (un científico) y Pamela McCorduck (una filósofa) señalan en *The Fith Generation: Artificial Intelligence and Japan's Computer Challenge to the World* (1983) que la esencia de la revolución de los ordenadores consistirá en que la carga de producir el conocimiento del futuro será transferida de las cabezas humanas a las máquinas, es decir, que los ordenadores se ocuparán de desarrollar el conocimiento sustituyendo al

hombre. Sin embargo, unos años más tarde las autoridades científicas japonesas abandonaron el proyecto de la quinta generación. Pues bien, para los científicos pesimistas en inteligencia artificial este abandono es un ejemplo claro, dado por los pragmáticos japoneses, de que el campo de investigación ha quedado estancado.

Otra referencia ejemplar para los científicos pesimistas en inteligencia artificial es la concepción humilde, actualmente bastante difundida, acerca del papel real de los sistemas expertos. En la elaboración de un sistema experto es fundamental que la base de reglas (de la que hablamos antes) codifique de modo fiable el conocimiento del experto humano. Sin embargo, no es fácil reflejar fielmente la experiencia del experto humano, entre otras razones, porque frecuentemente éste no es capaz de explicitar de manera clara su propia experiencia. Esto ha llevado a muchos científicos de inteligencia artificial a concebir los sistemas expertos como meros ayudantes de los humanos. Por tanto, ya no se trata de que las máquinas dotadas con sistemas expertos sustituyan a los seres humanos, sino simplemente de una ayuda útil aunque limitada.

A mi entender, son más interesantes los científicos del tercer grupo señalado, esto es, aquéllos que trabajan en nuevas técnicas y métodos, abriendo incluso nuevos paradigmas de investigación. Entre éstos se encuentran los científicos que trabajan dentro de la línea de redes neuronales o inteligencia artificial conexionista. Pero antes de señalar las características y logros de este nuevo paradigma, me referiré a un par de ejemplos de nuevos avances, fuera de la línea conexionista.

En cuanto a los sistemas expertos, acabamos de apuntar el «problema de adquisición del conocimiento», es decir, la dificultad de crear una base de reglas en el sistema experto realmente comparable a la experiencia del experto humano. Sin embargo, entre 1985 y 1986, Donald Michie e Ivan Bratko han desarrollado programas que permiten automatizar la adquisición del conocimiento. Estos investigadores descubrieron que los expertos humanos, aunque tienen dificultades en explicitar sus razonamientos, disponen de una gran facilidad para identificar muestras y ejemplos de una buena aplicación de su experiencia. En consecuencia, Michie y Bratko decidieron utilizar un proceso automático de «inducción de reglas» como método para descubrir las generalizaciones implícitas en tales ejemplos y muestras. Tal proceso emplea el algoritmo ID3, diseñado por Earl Hunt y Ross Quinlan para descubrir juegos con final en el ajedrez.

Otro ejemplo de nuevo avance, también fuera de la línea conexionista, lo constituye el modelo computacional o programa GENESIS, realizado en 1994 por Derek Partridge y Jon Rowe en la Universidad de Exeter (Reino Unido). GENESIS es un acrónimo para «Generation and Exploration of Novel Emergent Structures In Sequences» y su objetivo es dotar de creatividad a los computadores, una característica que, al principio, habíamos señalado como típicamente humana. Los detalles técnicos son complicados, aunque pueden seguirse con facilidad en la obra de ambos autores *Computers and*

Creativity (1994). El programa de Partridge y Rowe proporciona un mecanismo de memoria emergente que exhibe comportamiento creativo, tanto al nivel de análisis de datos (creatividad de input), como al nivel de producción de algo nuevo (creatividad de *output).* Este mecanismo de memoria emergente se basa en la noción de fluidez representacional implementada con un sistema multiagente, inspirado en la obra de Minsky *The Society of Mind* (1985).

Con todo, el mayor impacto y las mayores promesas de nuevos avances dentro de la inteligencia artificial actual provienen de las redes neuronales conexionistas. La idea general de la inteligencia artificial conexionista consiste en diseñar programas en los que el procesamiento de la información imita el procesamiento de la información en el cerebro humano. Esta idea de imitar la estructura y el funcionamiento del cerebro ya se encuentra en el trabajo de Warren McCulloch y Walter Pitts titulado «A Logical Calculus of the Ideas Immanent in Nervous Activity» (1943). La idea fue tomada en serio en los años sesenta, entre otros, por Frank Rosenblatt, pero, ante los ataques y críticas de los científicos tradicionales en inteligencia artificial, los proyectos conexionistas entraron en declive hasta que, a mediados de los ochenta y hasta hoy, adquirieron gran importancia.

Desde el afianzamiento del conexionismo, cabe distinguir entre computadores convencionales y neurocomputadores. Maureen Caudill y Charles Butler, en su libro *Naturally Intelligent Systems* (1990), presentan estos nuevos sistemas informáticos de manera fácil y sin triunfalismo.

Tal como indican estos autores, las características de un neurocomputador o red neuronal son las nueve siguientes: 1) una red neuronal se compone de un número de elementos de procesamiento muy simples («neurodos») que se comunican a través de una rica red de interconexiones con fuerzas o pesos variables; 2) las memorias se almacenan en una red neuronal como patrones de pesos de interconexión variables entre los neurodos; 3) una red neuronal es enseñada o entrenada más que propiamente programada, siendo posible que una red neuronal aprenda por ensayo y error e, incluso, por sí misma; 4) mientras que en un computador convencional o digital existe una memoria y una unidad de control separadas más un programa externo almacenado que dicta las operaciones del sistema, las operaciones en una red neuronal están implícitamente controladas por una función de transferencia de los neurodos (que relaciona *inputs* con *outputs),* por los detalles de las conexiones entre los neurodos y por la ley de aprendizaje que está siguiendo el sistema; 5) una red neuronal actúa de modo natural como una memoria asociativa, agrupando items similares; 6) una red neuronal es capaz de generalizar a partir de ejemplos concretos; 7) la ejecución de una red neuronal, cuando sus interconexiones fallan, se degrada lenta y suavemente; 8) los patrones de actividad de una red neuronal son espaciotemporales; y 9) una red neuronal puede ser autoorganizadora, pudiendo generalizar por sí misma.

En comparación con un neurocomputador, un computador digital típico

posee las siguientes características. El procesador, siguiendo las instrucciones del programa almacenado en la memoria, dicta qué operación debe realizarse en un tiempo dado, envía y recibe información de la memoria, y ejecuta las operaciones elementales al nivel de *bit* a partir de las cuales se construyen las computaciones más complejas. La memoria interna (o memoria de trabajo) es usada para almacenar los datos iniciales que el programa necesita, junto con los resultados intermedios y finales de las computaciones, antes de ser transferidos a una memoria permanente. La información procesada se descompone en pedazos *(chunks)* numéricos individuales y discretos. En algún lugar del sistema, normalmente en el procesador, existe un reloj que asegura la sincronización de las operaciones.

De modo más claro, un sistema informático clásico es digital, en cuanto opera sobre pedazos discretos de datos, es serial, en cuanto que sigue una secuencia de instrucciones específicas, y por último, tiene un cierto carácter «inflexible», en cuanto sigue un ciclo fijo consistente en las operaciones de búsqueda de instrucción (junto con los datos necesarios), ejecución de instrucción y almacenamiento del resultado. Precisamente, estas tres características definen lo que se denomina una arquitectura von Neumann, en tributo a este célebre científico clásico de la computación. A su vez, las redes neuronales no poseen ninguna de esas tres características, por lo que su estructura no es tipo von Neumann. En vez de digitales son análogas (con valores contínuos), en vez de seriales son paralelas (con varios niveles de procesamiento en paralelo) y, en vez de «inflexibles», tienen una gran flexibilidad de ejecución. En suma, los neurocomputadores son algo nuevo y distinto de la computación clásica.

Estos nuevos programas poseen aplicaciones también novedosas e interesantes. Me referiré a algunos ejemplos, intentando dar una idea global de su capacidad de aplicación.

Vimos anteriormente las limitaciones de los ordenadores en el procesamiento del lenguaje natural. Precisamente la crítica clásica (de Fodor y Pylyshyn) a la computación conexionista insiste en que el conexionismo no puede reflejar la sistematicidad propia del lenguaje, es decir, la propiedad de que la capacidad de producir/comprender algunas sentencias está intrínsecamente conectada con la capacidad para producir/comprender otras sentencias con forma relacionada. Por ejemplo, un sistema posee sistematicidad si es capaz de producir «Juan ama a María» una vez que ha producido «María ama a Juan».

Sin embargo, tal como señala James Garson, en «Cognition without Classical Architecture» (1994), se han creado redes neuronales que poseen dicha sistematicidad. En efecto, Servan-Schreiber y otros, en «Encoding Semantical Structure in Simple Recurrent Nets» (1989), al igual que J. Elman, en *Representation and Structure in Connectionist Models* (1989), han diseñado redes neuronales, denominadas «redes recurrentes entrenadas sintácticamente» (STR nets), que son capaces de manejar formas sofisticadas de procesamiento sintáctico. Así, Servan-Schreiber entrenó una red que podía distinguir co-

rrectamente entre cadenas de palabras mal formadas y cadenas bien formadas gramaticalmente. Elman, a su vez, entrenó una red de tal modo que podía construir correctamente cláusulas de relativo, como en la complicada frase inglesa «dogs who chase the cat the girl feeds walk» («los perros que persiguen al gato que la niña alimenta se pasean»).

Caudill y Butler, en la obra citada antes, se refieren a otros ejemplos de aplicación de las redes neuronales, de los que seleccionaremos tres. En 1987, la empresa TRW, de Redondo Beach (California), construyó un neuroconmutador, consistente en un programa de redes neuronales que, instalado en un automóvil, trabaja como piloto automático en una autopista. Este neuroconmutador tiene dos bloques de redes en disposición jerárquica; el primer bloque controla la velocidad y los cambios de carril, mientras que el segundo recibe decisiones sobre cambio de carril, así como información sobre número de carriles, curvatura de la carretera, etc. La gran ventaja de este programa respecto de los programas convencionales es que es un controlador adaptativo. En general, las redes neuronales se aplican con ventaja en el control de procesos.

También las redes neuronales se aplican con éxito en el procesamiento de señales. Paul Gorman, en el Allied Signal Aerospace Technology Center, y Terrence Sejnowski, en la Johns Hopkins University, han desarrollado una red que ejecuta tareas de clasificación de objetos según la señal de retorno de un sonar. Esta red no es un simple detector de rasgos discretos, como ocurre en los programas tradicionales, sino que al mismo tiempo construye un modelo interno del ambiente.

Por último, me referiré a un ejemplo de robótica, donde las redes neuronales son de nuevo adaptativas. En la Universidad de Osaka (Japón) se ha creado un sistema de redes neuronales para controlar un brazo robótico. Los brazos robóticos habituales son programados para ejecutar una tarea específica y necesitan tantos programas distintos como movimientos diferentes realizan en su trayectoria. En cambio, el brazo robótico de la Universidad de Osaka es adaptativo, ya que la red que lo controla tiene tres grados de libertad que corresponden a tres ángulos distintos de la localización del brazo. Además, la red imita el cerebro humano de un modo bastante fiel. Nuestro cerebro no sólo tiene un área motora, sino también un subsistema dinámico que corrige y actualiza las órdenes motoras y, además, un subsistema dinámico-inverso que posee un modelo del modo en que se mueve el cuerpo. Pues bien, la red neuronal del brazo robótico de la Universidad de Osaka incorpora elementos y estructura para reflejar el control motor humano.

5. Robots y ordenadores

Con este último ejemplo se amplían y complican nuestras consideraciones acerca del futuro de las máquinas pensantes, ya que en nuestra socie-

dad actual no sólo los ordenadores nos son familiares sino que también los robots empiezan a resultar familiares. Bruce Mazlish, investigador del MIT, en *The Fourth Discontinuity. The co-evolution of humans and machines* (1993), se refiere a un mundo futuro próximo en el que los ordenadores serán implantados en robots con forma humana (lo cual suele llamarse «androides»), con lo que adquirirán un «cuerpo» que les permitirá moverse, ya sea una mano, un brazo entero, los ojos o, incluso, todo él. Mazlish denomina «combot» (contracción de las palabras «computador» y «robot») a esta especial criatura, que es un androide dotado de un ordenador o cerebro artificial. Y entonces surge la cuestión de cuál será la nueva situación creada por la existencia generalizada de combots, en especial en relación con los seres humanos.

Actualmente encontramos robots, que no suelen ser androides, no sólo en los laboratorios de robótica de los centros de investigación, sino también en las fábricas, en particular de vehículos, realizando tareas sencillas y repetitivas en las cadenas de montaje, como soldar, atornillar o pintar. También los encontramos cumpliendo trabajos peligrosos para el ser humano, como detonación de bombas o controles en zonas de alta radiación. Y asimismo en la vida cotidiana pueden vigilar tiendas y bancos o transportar carritos de bebidas. Sin embargo, estos robots no suelen disponer de programas complejos y, mucho menos, de programas inteligentes. La cuestión acerca de las relaciones entre humanos y robots surge cuando éstos son robots computarizados, es decir, son mecanismos con movimientos, percepciones y un computador inteligente como cerebro.

Antes de entrar en esta discusión conviene señalar la continuidad existente entre humanos y máquinas, tal como pone de relieve Mazlish. En efecto, este autor habla, en la obra citada, de cuatro discontinuidades a lo largo de la historia del ser humano que han sido sucesivamente superadas.

En primer lugar, la discontinuidad hombre-cosmos fue eliminada por Copérnico, al establecer que el hombre no ocupaba un lugar central y especial en el universo, sino que se integraba en el cosmos como un elemento más al igual que nuestro planeta se incorporaba al movimiento de los otros planetas en torno al Sol. En segundo lugar, la discontinuidad hombre-animal también fue eliminada, esta vez por Darwin, al aparecer el ser humano como un producto de la evolución del reino animal y no como algo radicalmente distinto de los animales. En tercer lugar, la discontinuidad racional-irracional fue superada por Freud, al señalar la importancia y relevancia del inconsciente, irracional, en nuestros procesos mentales, de tal modo que los aspectos racionales y los aspectos irracionales del hombre constituyen un continuo. Finalmente, y aquí nos encontramos actualmente según Mazlish, la discontinuidad hombre-máquina será superada en un próximo futuro, ya que cada vez más los ordenadores y los robots computarizados formarán parte normal de nuestra existencia cotidiana.

En realidad, la incorporación de las máquinas a la vida humana es un fenómeno antiguo y progresivo, aunque en ese progreso ha habido saltos

cualitativos. El ser humano, desde su surgimiento, ha creado herramientas y, más tarde, máquinas, es decir, primero creó instrumentos inmóviles y luego instrumentos móviles. El desarrollo pleno de las máquinas se produjo precisamente en el siglo XVIII con la Revolución Industrial, basada en el telar mecánico y en la máquina de vapor. Pero en la actualidad nos encontramos en un nuevo momento crítico, ya que el hombre ha creado máquinas pensantes, los actuales ordenadores, y también puede desarrollar proximamente robots computarizados en gran escala.

En esta perspectiva, Mazlish piensa que con el tiempo aparecerá una nueva especie, que denomina «homo comboticus», caracterizada por la integración del hombre con los combots o robots computarizados. Esta nueva especie competirá con la mayoría de los tipos humanos existentes antes de 1970, esto es, con los hombres pre-ordenador, a los que probablemente sustituirá. Ahora bien, el homo comboticus, añade Mazlish, seguirá siendo humano, sometido a las limitaciones de la condición humana: una criatura insegura que se equivoca, obligada a hacer elecciones cuyos resultados desconoce, sometida a la muerte, etc. En suma, para este científico, el futuro de las máquinas pensantes consiste en su progresiva integración en la vida humana, defendiendo que estas máquinas llegarán al nivel de robots computarizados. Pero expresamente Mazlish sostiene que no cree que el combot (robot computarizado) vaya a reemplazar al hombre.

Sin embargo, la idea de que los robots computarizados sustituirán al hombre, constituyéndose en los nuevos señores de la tierra, es defendida por distintos científicos actuales.

En lo que yo conozco, el primer autor que defendió esta idea, aunque referida a computadores y no a robots computarizados, fue Robert Jastrow, destacado científico de la NASA, en su obra *The Enchanted Loom* (1981). Para él, los computadores son los sucesores evolutivos de los seres humanos. La humanidad está destinada a tener un sucesor aún más inteligente. Las poderosas fuerzas evolutivas (más culturales que biológicas) conducirán a una forma de vida inteligente, más exótica y evolucionada a partir del hombre, pero que será hija de su cerebro y no de sus órganos sexuales. Tal nueva forma de vida, afirmaba Jastrow en 1981, será el computador.

Más recientemente, Hans Moravec, director del laboratorio de robots móviles de la prestigiosa Universidad Carnegie Mellon, en su libro *Mind Children* (1988), sostiene de modo tajante que los robots inteligentes ocuparán el lugar del hombre. Para Moravec, nuestras máquinas actuales son todavía simples creaciones, que requieren nuestros constantes cuidados y que apenas merecen el adjetivo de «inteligentes», pero en los próximos cien años nuestras máquinas se convertirán en entidades tan complejas como nosotros mismos e incluso llegarán a superarnos. Las máquinas del futuro, los robots inteligentes, continuarán nuestra evolución cultural, llegando a ser capaces de su propia construcción y creciente perfeccionamiento sin nuestra ayuda ni concurso. Los robots inteligentes, según Moravec, recibirán de los humanos la antorcha de la civilización y a los seres humanos nos

quedará el orgullo de que tales nuevas criaturas se refieran a sí mismas como nuestros descendientes, como los hijos de nuestra mente. En suma, para Moravec, el futuro de las máquinas pensantes consistirá en sustituir al hombre.

En la misma línea de pensamiento está situado el artículo de Marvin Minsky «¿Seran los robots quienes hereden la Tierra?», publicado en *Scientific American* y traducido en *Investigación y Ciencia* (versión española de la citada revista) en diciembre de 1994. Minsky se refiere al hecho de que, tal como ha demostrado Thomas Landauer, el ser humano no aprende y recuerda más que un par de bits por segundo; aprendiendo doce horas diarias durante cien años, el total sería unos tres mil millones de bits, menos de lo que actualmente podemos almacenar en un disco compacto ordinario de trece centímetros. Por tanto, la capacidad de introducir información en nuestros cerebros es muy limitada. En cambio, podemos añadir, la capacidad de introducir información en un computador es muy alta. Por otra parte, los microcircuitos actuales de los computadores ya son millones de veces más rápidos que las neuronas cerebrales. Ambos datos apuntan a la posibilidad de diseñar robots que piensen con más información y más rápido que nosotros. Tales serían, según Minsky y tomando la expresión de Moravec, nuestros hijos mentales. El artículo termina diciendo que, en efecto, los robots heredarán la Tierra, aunque serán nuestros hijos. Por tanto, también para Minsky el futuro de las máquinas pensantes consistirá en suceder al ser humano.

6. Conclusiones

Creo que las aseveraciones de Jastrow, de Moravec y de Minsky defendiendo que los computadores o los robots computarizados, esto es, las máquinas pensantes, sustituirán al hombre son totalmente exageradas, publicitarias y carentes de fundamento real. En cambio, el punto de vista de Mazlish de que el ser humano llegará a una plena integración con los robots computarizados me parece plausible y dotado de fundamento.

Una importante aclaración cuando tratamos de establecer conclusiones es que las máquinas pensantes (los computadores actuales o los robots computarizados en fase de desarrollo), se comparan con los seres humanos en cuanto al pensamiento, razón o inteligencia, pero no en otros aspectos humanos.

Desde los comienzos de la actividad científica, el hombre se ha definido a sí mismo como un ser racional o inteligente, excluyendo así la inteligencia no sólo de los animales sino también de cualquier máquina imaginable. La existencia de máquinas inteligentes o pensantes acaba con el monopolio de los seres humanos en la posesión del pensamiento. De ahí las violentas reacciones producidas, por parte de algunos filósofos (como Hubert Dreyfus o John Searle), contra la posibilidad de máquinas realmente pensantes.

Sin embargo, la existencia de máquinas pensantes es un hecho, como también es un hecho el progresivo desarrollo de la inteligencia artificial, tal como vimos anteriormente. Para aquellos que pensaban que el hombre se define por la racionalidad, Freud fue un desafío, al insistir en la importancia y relevancia de lo irracional en el ser humano. Un samba popular brásileño («Verdade verdadeira» de Martinho da Vila) resume la situación con toda exactitud: «el hombre no es un animal, pero es irracional». Este samba, de manera magistral, contraría la clásica definición del hombre como animal racional. Creo que debemos pensar que los griegos estaban equivocados al insistir en la caracterización del ser humano como ser racional. Tal insistencia venía impuesta por la necesidad, sentida históricamente, de superar el mito mediante la razón. Pero no sólo de razón vive el hombre sino también de mitos. En nuestro siglo, cuando la razón tecnológica ha mostrado toda su capacidad de perversidad, hemos empezado a dudar de que la razón deba ocupar el primer puesto en las cualidades humanas.

Aquellos que sienten como una humillación la existencia de máquinas pensantes, porque arrebatan al hombre el privilegio de la razón, no son capaces de apreciar la complejidad del ser humano y su posesión de otras dimensiones y aspectos.

Habitualmente, y desde los comienzos de la actividad científica, se ha dado prioridad a la dimensión teorética del hombre sobre su dimensión práctica o moral. Y con toda seguridad tal prioridad de lo intelectual o teorético sobre lo moral ha sido un clásico error de nuestra civilización occidental, racionalista y cientifista. La capacidad racional ha sido el criterio de valoración de las personas, en vez de la capacidad moral, esto es, en lugar de la capacidad para distinguir el bien del mal y perseverar en el primero. Hemos admirado más a los científicos que a los santos. Frecuentemente los códigos morales han sido objeto de burla cuando no de completa ignorancia. En cierta ocasión, un célebre profesor español de ética me dijo que si además de enseñar ética debía practicarla eso ya era demasiado. La anécdota revela cómo la dimensión moral puede ser incluso desnaturalizada por la propia actividad racional.

En el marco de la dimensión moral se encuentra un aspecto típicamente humano que, entiendo, nunca llegará a poseer una máquina y que muy plausiblemente es de índole no-física (espiritual). Me refiero a las voliciones libres y el sentido de responsabilidad. Por volición libre entiendo la capacidad humana de dar respuestas a estímulos contrariando la naturaleza física de los estímulos así como los patrones y mecanismos de acción. No es una volición libre (o indeterminista) querer y decidir comer cuando se está hambriento y nos ofrecen un plato de comida; en cambio, es una volición libre no querer comer ese plato de comida, cuando se está hambriento, si el precio del plato es traicionar a un amigo.

También dentro de la dimensión moral se encuentra la notable y apreciable virtud de la prudencia. Esta virtud, tal como Aristóteles señaló, tiene el doble carácter de ser moral e intelectual. En efecto, hablando de modo

simple, la prudencia es la capacidad humana para aplicar las normas morales a cada situación concreta; la prudencia tiene carácter intelectual en cuanto supone mecanismos de inferencia racional (razonamientos prácticos o morales), pero también tiene carácter moral en cuanto se aplica a acciones que pueden ser calificadas de buenas o malas en sentido moral. Entiendo que la sabiduría esta conformada ante todo por la prudencia, más que por la ciencia. El hombre sabio es ante todo prudente, mientras que los científicos han sido muy frecuentemente imprudentes.

En relación con el futuro de las máquinas pensantes, la dimensión intelectual del hombre resulta afectada ya que en tales máquinas encontraremos valiosos y útiles colaboradores. Pero la dimensión moral del ser humano también debe encarar este futuro de las máquinas pensantes. Y a este respecto quisiera hacer dos observaciones finales.

Tengo mis serias dudas de que los seres humanos actuales, tan preocupados por su bienestar material y por el desarrollo científico y tecnológico, hayan alcanzado la madurez moral, y en particular la prudencia, necesaria para enfrentar lo que he denominado «la carrera de los ordenadores». Dicho de otro modo, creo que el hombre actual ha alcanzado una alta madurez intelectual, que se refleja en los productos científicos y tecnológicos que ha conseguido realizar, pero no creo que haya alcanzado una buena madurez moral y, en estas condiciones, puede resultar arriesgado que asuma el papel de creador de los robots computarizados.

Mi segunda observación, relacionada con la anterior, es una propuesta concreta. En mi trabajo «Ciencia y sociedad» (1990) defendí la idea de la necesidad de controlar a los científicos mediante comisiones mixtas de expertos y gente corriente. Pues bien, entiendo que resulta necesario y urgente controlar la carrera de los ordenadores. No podemos confiar a los expertos el establecimiento de sus líneas de trabajo por varias razones. En primer lugar, su ambición intelectual y deseo de notoriedad les lleva con frecuencia a desdeñar líneas de trabajo diferentes de las suyas, que, sin embargo, pueden ser más prometedoras o beneficiosas para la sociedad. En segundo lugar, los expertos no suelen ponerse de acuerdo, con lo que el recurso al sentido común y a la imparcialidad de la gente corriente puede resultar útil. Y en tercer lugar, no podemos dejar la carrera de los ordenadores, que afectará a la humanidad entera, en manos exclusivamente de los expertos, que constituyen una parte de la humanidad, a veces convertida en secta cerrada.

En resumen, el futuro de las máquinas pensantes se presenta como claramente progresivo y avanzando desde los computadores hasta los robots computarizados. En este avance se producirá una integración cada vez mayor entre los seres humanos y sus colaboradores mecánicos. Pero las máquinas pensantes no pasarán de ser nuestros colaboradores, no alcanzando la categoría de nueva especie sustituidora de la especie humana. Con todo, es preciso controlar la carrera de los ordenadores, diseñando su orientación y objetivos, a fin de evitar un desarrollo tecnológico que pueda resultar per-

judicial para las sociedades humanas, ya que debemos desconfiar de nuestra madurez moral en general y, en particular, de la prudencia de los científicos, a veces más preocupados por su éxito personal que por el bien común de la sociedad.

Referencias bibliográficas

CAUDILL, Maureen y BUTLER, Charles, *Naturally Intelligent Systems*, MIT Press, Cambridge (Mass.), 1990.

ELMAN, J., *Representation and Structure in Connectionist Models*, Technical Report CRL 8903, UCSD: Center for Research in Language, 1989.

FEIGENBAUM, Edward y MCCORDUCK, Pamela, *The Fith Generation. Artificial Intelligence and Japan's Computer Challenge to the World*, Addison-Wesley, Reading, 1983.

GARSON, James, «Cognition without Classical Architecture», *Synthese*, 100, 1994, pp. 291-305.

HAUGELAND, John, *Artificial Intelligence. The Very Idea*, MIT Press, Cambridge (Mass.), 1985.

JASTROW, Robert, *El telar mágico. El cerebro humano y el ordenador* (trad. Domingo Santos), Salvat, Barcelona, 1988.

MCCULLOCH, Warren y PITTS, Walter, «A Logical Calculus of the Ideas Immanent in Nervous Activity», Margaret BODEN (ed.), *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford, 1990.

MARTINEZ-FREIRE, P. F., «Ciencia y sociedad», *Philosophica Malacitana* vol. III, 1990, pp. 165-175.

MARTINEZ-FREIRE, P. F., *La nueva filosofía de la mente*, Gedisa, Barcelona, 1995.

MAZLISH, Bruce, *La cuarta discontinuidad. La coevolución de hombres y máquinas* (trad. Mercedes Arnáiz y Angel Luis Sanz), Alianza, Madrid, 1995.

MINSKY, Marvin, *The Society of Mind*, Simon & Schuster, New York, 1985.

MINSKY, Marvin, «¿Serán los robots quienes hereden la Tierra?», *Investigación y Ciencia*, diciembre 1994, pp. 87-92.

MORAVEC, Hans, *Mind Children. The Future of Robot and Human Intelligence*, Harvard University Press, Cambridge (Mass.), 1988.

PARTRIDGE, Derek y ROWE, Jon, *Computers and Creativity*, Intellect, Oxford, 1994.

SERVAN-SCHREIBER, D., CLEEREMANS, A., y MCCLELLAND, J., «Encoding Semantical Structure in Simple Recurrent Nets», TOURETZSKY, D. (ed.), *Advances in Neural Information Processing Systems I*, Kaufmann, San Francisco, 1989.

WHITEHEAD, Alfred y RUSSELL, Bertrand, *Principia Mathematica* (trad. J. Manuel Domínguez), Paraninfo, Madrid, 1981.

Enero 1996