ALGUNAS OBSERVACIONES SOBRE NARCISISMO HUMANO E INTELIGENCIA ARTIFICIAL[1]

Alejandro Sobrino

Área de Lógica y Filosofía de la Ciencia. Campus Universitario, s.n.
15706 Santiago de Compostela

En 1917, Freud señaló que el narcisismo humano había sido ofendido en tres ocasiones: la ofensa cosmológica, cuando Copérnico descubre la dependencia cinética de la Tierra respecto al Sol; la ofensa biológica, proviniente de los trabajos de Darwin que sitúan al homo como una especie entre otras y, por último, la que resulta de su propia teoría psicoanalítica: la ofensa que el subconsciente infringe a una conciencia que se cree dominadora de la trastienda humana. Desde hace algún tiempo, es un tema polémico si la Inteligencia Artificial y, en concreto, un exitoso desarrollo suyo, los Sistemas Basados en el Conocimiento, constituyen también una ofensa al narcisismo humano, en tanto que desmitifican alguna parcela tradicionalmente considerada como especie-específica: la facultad razonadora. Al respecto se han vertido afirmaciones exageradas desde distintos ámbitos de las Ciencias Cognitivas: desde la Filosofía o la Ingeniería del Conocimiento. En este trabajo se propone un marco teórico en el que debatir con sentido estas cuestiones con algunos criterios que creemos más racionales y útiles que los hasta ahora empleados.

Palabras clave: narcisismo humano, inteligencia artificial, sistemas basados en el conocimiento, especie-específica, facultad razonadora.

1. Introducción

Una característica que comúnmemente admitimos para los homo como género natural diferenciado es la de poseer capacidad simbólica y, más especificamente, habilidad lingüística. De siempre, el lenguaje ha sido considerado un vehículo apropiado y útil para depositar y transmitir

1 Financiación a cargo de los proyectos 2050B90 de la Xunta de Galicia y Ps89-0175 de la D.G.I.C.Y.T.

nuestros pensamientos, si el pensamiento precede al lenguaje o si ambos van al unísono. El lenguaje le ha servido al homo, no sólo para hablar de otros entes conocidos, sino, sobre todo, para referirse a si mismo. Esta capacidad autorreferencial se le ha denominado usualmente consciencia y se la ha situado, con frecuencia, en la raiz de la hominidad.

La conciencia de si mismo ha permitido a los humanos desarrollar, por lo menos, dos tipos de comparaciones.

(a) Comparaciones con otros seres.

Habitualmente el homo se ha comparado con entes extraños a si mismo en una doble modalidad: (a1) Con seres que poseen su misma condición de 'ser vivos'; esto es, con otros animales. (a2) Con artefactos ajenos a su ser, pero que llevan el cuño de su manufactura. Desde la aparición de la Inteligencia Artificial (IA), este tipo de comparación ha arreciado con fuerza, por lo que hemos considerado conveniente discutir en este trabajo algunos aspectos relacionados con la misma.

(b) Comparaciones entre distintas modalidades de su ser.

La consciencia de si mismo posibilita que el homo se catalogue en virtud de sus propias características, como el color de la piel, las costumbres o el hábitat, en una tipología que sin duda es interesante para la biología o la antropología. Pero la autorreferencia permite una distinción más relevante desde el punto de vista de este estudio: el que el cerebro humano se escrute a si mismo posibilita diferenciar las acciones humanas en mentales y corporales. El interés de esta distinción es que, con frecuencia, ha servido para caracterizar a los fenómenos mentales como aquellos que son irreductiblemente humanos.

Desde el punto de vista de la IA, la distinción entre mente-cerebro, aunque muy utilizada en el ámbito de las ciencias cognitivas más especulativas --como la filosofía de la mente, no parece excesivamente clarificadora para evaluar el alcance de algunas afirmaciones acerca de las incapacidades o posibilidades del proyecto IA. El objetivo de este trabajo es proponer algunas distinciones más explicativas para este propósito. Nues-

tra pretensión, por tanto, es en cierto modo propedéutica: el propósito es dar algunas claves para leer mejor algunas tesis sobre las afinidades o diferencias entre el homo y la máquina electrónica. No se pretende, en ningún caso, dar un marco general de discusión acerca de este tema.

2. Ofensas al narcisismo humano

En su comparación con otros seres vivos, el homo desequilibró muy pronto la balanza a su favor al atribuirse los títulos de 'Centro del Universo' y 'Rey de la Creación'. El homo se siente 'Rey de la Creación' por una suerte de confirmación empírica: ya que en la mayoría de los casos hace favorables a su causa las situaciones o acciones en las que intervienen otros seres, debe considerarse a si mismo como una especie distinta y superior al resto. La certificación del primer título, 'Centro del Universo', era concedida por las teorías físicas que otorgaban a la Tierra, su vivienda, el estatuto de pentagrama en función del cual se escribía la solfa del Universo. Ambos atributos eran apoyados además por otros dos factores, sin duda extraños, pero importantes para afianzar su autoestima: uno, la ausencia de vestigios de vida extraterrestre (de competidores); otro, las numerosas prédicas religiosas que situaban al homo como especie elegida por Dios como reflejo de si mismo. El homo legitimaba así con una construcción simbólica (la religión) otra construcción simbólica (su puesto en el Cosmos) y alimentaba el orgullo de ser una especie única y preferente.

Fue la investigación científica la que restringió progresivamente parcelas a este sentimiento de omnipotencia humano. En un ensayo que cosntituye no sólo una aportación al psicoanálisis, sino también un bello ejercicio literario, Freud señaló que, por lo menos, en tres ocasiones había sido cuestionado nuestro narcisismo (Freud, 1917: 2434-2435). Según la distinción introducida antes, las dos primeras ofensas surgieron de la comparación con otros seres vivos; la tercera se apoyó en la distinción freudiana de dos modalidades de nuestro 'ser': el ser consciente y el ser inconsciente.

(i) A la primera ofensa Freud la denomina la *ofensa cosmológica* y se apoya en la demostración de Copérnico, en el siglo XVII, del movimiento de la Tierra en torno al Sol. La Tierra pasa a depender cinéticamente de

otro planeta y la casa del homo ya no ocupa un lugar central en el Cosmos conocido.

(ii) De la segunda dice que es la *ofensa biológica*; el ofensor es Darwin con sus trabajos sobre la evolución de las especies que, a partir de entonces, sitúan al homo no como un inquilino preferente del solar cósmico, sino simplemente, como ocupando un celdaño en la escala zoológica, emparentado con más proximidad a unos seres que a otros, pero, en cualquier caso, heredando siempre caracteres biológicos de especies anteriores y más primitivas que él.

(iii) La tercera ofensa lleva el nombre de *ofensa psicológica*, y está asentada en la distinción psicoanalítica entre 'ser consciente' y 'ser inconsciente'. Como es conocido, según la teoría psicoanalítica, en numerosos actos de humanos hay un dominio de lo inconsciente sobre lo consciente, de la pasión sobre la razón. El homo declara su intención de ejercitar lo razonable, pero es en buena medida instinto y pasión y un examen de cualquier 'conducta desviada' así lo muestra. En estos casos, lo anímico no coincide con lo consciente, la pulsión es distinta a la razón, '*el yo no es dueño y señor de su propia casa*', sino que está a expensas de lo que su trastienda le ordene. Esto compromete muy seriamente al narcisismo humano ya que, si bien el homo había sido '*exteriormente humillado*, (no obstante, hasta entonces) *se sentía soberano de su propia alma*' (Freud, 1917: 2436).

La ofensa psicológica es importante. Si el homo no es señor de su propio yo, tampoco puede aspirar a ser 'el ser' entre los seres o a que su vivienda, la Tierra, ocupe un lugar central en el Universo. Pero hay otro tipo de ofensa, que también surge de la psicología y que es especialmente interesante para el tema que queremos estudiar.

Como ya adelantamos, la agresión surge, en este caso, de la comparación del homo con productos tecnológicos que el mismo trata de desarrollar. Sin embargo, tiene presencia en un ámbito de la psicología distinto al psicoanálisis: se trata de la psicología cognitiva, cuyo objeto de estudio son los procesos cognitivos humanos, como p. ej., la memoria o el procesamiento de información. Es todavía una ofensa provisional, ya que no se basa en una realización concreta, sino en un gran proyecto científico contemporáneo: la construcción de artefactos que simulen (o reproduzcan, como tendenciosamente le gusta decir a Hal, el ordenador de 2001)

algún aspecto de la inteligencia humana, con un porcentaje de éxitos elevados respecto a aquellas tareas para las cuales se haya propuesto la simulación.

El proyecto de las Ciencias Cognitivas (a las que presta su apellido la Psicología Cognitiva) constituye una ofensa al narcisismo humano porque en él se observa habitualmente:

(a) La reducción de fenómenos psicológicos a fenómenos físicos, en el proceso puesto en marcha por la neurobiología.

En efecto, si el descubrimiento de aspectos subconscientes en conductas conscientes supone una ofensa al narcisismo del hombre dueño de si mismo, la reducción (explicación) de aspectos considerados como típicamente humanos (subconscientes, irracionales) a su correlato físico, supone quitarle trocitos a esa inmensa tarta que se supone es nuestra humanidad. Los neurobiólogos, p. ej., mostraron que la epilepsia, considerada hasta hace poco como una enfermedad mental, se caracteriza hoy en día bastante bien en términos fisiológicos. El método de las ciencias físicas, tradicionales estudiosas de objetos inertes, explicaría así un fenómeno irracional, propio de un ser vivo, en términos de la metodología con la que se estudian los seres no vivos.

(b) La separación de las conductas intimistas de aquellas que pueden ser objetivables intersubjetivamente, promoviendo una psicología no introspectiva donde, aspectos secularmente considerados como 'problemas del alma', reciben algún tipo de explicación científica.

La diferenciación entre lo introspectivo y lo objetivable ha tenido repercusiones positivas, no sólo en la demarcación del ámbito de estudio de algunas ciencias ligadas a la IA, como la lógica y la psicología, sino también para dirimir con exactitud el objeto de estudio de la propia IA. Así, con Frege, la lógica se constituyó en una ciencia que estudia la capacidad argumentativa o razonadora del ser humano, no como proceso psíquico, sino como resultado de alguna proferencia o grafía. El destierro del introspeccionismo de la lógica ha sido un paso fundamental en el proceso de su consolidación como ciencia. Este mismo efecto lo ha surtido en la psicología que, de modo mayoritario, aspira hoy a dar un conjunto de conocimientos objetivables acerca de aspectos psíquicos del ser humano y a no ser simplemente un recetario intimista. La IA, deudora, entre

otros, de estos dos saberes, se ocupa de simular científicamente aquellos procesos que se pueden caracterizar objetivablemente como inteligentes.

3. Una primera distinción

Normalmente, se ha considerado en el homo como capacidad típicamente inteligente a su capacidad racional, su facultad para simbolizar cosas y hacer razonamiento con esos símbolos (normalmente lingüísticos) prescindiendo de los objetos. De ahí que al homo se le dé también a veces el atributo especie-específico de 'ser racional'. A la racionalidad del homo se le ha atribuído normalmente la marca de 'mental'. Creemos que la distinción mental/físico, aunque habitual en la literatura al uso, introduce confusiones respecto a lo que una buena mayoría de ingenieros de conocimiento --sobre todo aquellos que se ocupan de hacer predicciones a corto plazo-- cree que es su actividad científica. Aunque esta distinción es interesante para discutir algunos problemas teóricos, desde la perspectiva de la Ingeniería del Conocimiento parace conveniente proponer una alternativa.

En vez de la dicotomía mental/físico, que habitualmente se ha utilizado para señalar la separación entre aquello que es típicamente humano y aquello que, aún siendo humano, se puede explicar con el método de las ciencias naturales, en el ámbito de la IA parece adecuado desdoblar el carácter 'mental' en dos variedades: racional e irracional --o, en términos similares, mental objetivable y mental subjetivo. Esta distinción es evidentemente discutible, pero, en el contexto de este trabajo, aspiramos a mostrar que es útil. Por racional se entiende normalmente aquel comportamiento (normalmente lingüístico) guiado por algún tipo de juicio o razón. De un modo parcial y limitado, puede decirse que un comportamiento tal puede ser expresado, casi siempre simplificadoramente, por medio de reglas. Este tipo de racionalidad es la que tratan de simular los Sistemas Basados en el Conocimiento (SBC), una de las ramas más asentadas de la IA. Lo irracional, en cambio, es aquello que no se somete a reglas. Si bien puede considerarse que la capacidad racional y razonadora del homo es lo típicamente humano, lo conscientemente humano; lo irracional, expresión del subconsciente, podría ser caracterizado como lo 'más humano', lo irreductiblemente humano.

Como ya indicamos, la autorreferencialidad le permite al homo clasificar sus actos en mentales y corporales, de manera que a las facultades racionales, como razonar, se le atribuye frecuentemente el calificativo de 'mental'. Pero desde la IA no es carente de importancia distinguir entre un un 'acto mental' y el resultado de ese acto mental. Un acto mental puede ser mi disposición a hacer una inferencia; el resultado de ese acto es un argumento plasmado en algún tipo de proferencia o texto. Lo que interesa desde el punto de vista de la IA es lo último, no lo primero, en tanto que sólo los argumentos pronunciados o escritos pueden abandonar el reino de la intimidad para pasar a ser tema de análisis objetivable y, por tanto, útilmente simulable. Esta idea guía, en buena medida, los objetivos en IA, cuyos profesionales están más preocupados por las enormes dificultades que se encuentran en la simulación de conocimientos sumamente simples, que en hacer grandes afirmaciones de carácter predictivo. (Aunque son éstas las que más discusión alcanzan).

No obstante, a pesar de que el interés de la IA se centra en la simulación de aspectos cognitivos objetivables, puede plantearse también la siguiente cuestión: si se simula bien un comportamiento mental ¿en qué medida se está simulando también a la misma mente? Esta pregunta es frecuentemente debatida en los estudios especulativos sobre la IA, más que en ámbitos teóricos o aplicados de este proyecto, y alimenta muchos trabajos en filosofía de la mente. La cuestión en torno a si simulando un comportamiento mental se simula el acto mental previo a ese comportamiento es, sin duda, interesante desde un punto de vista teórico, pero sólo dirimible desde la filosofía de la IA, no desde la IA misma. El motivo se encuentra en que la respuesta forma parte del corpus de creencias del individuo que responde a tal cuestión, creencias que, en principio, no son objetivables. Esto no puede parecer extraño, ya que no ocurre sólo en la IA, sino que es habitual en otros saberes. P. ej., es un tema de la filosofía de la matemática discutir si cuando se demuestra un teorema se está descubriendo una verdad nueva o simplemente se desvela una verdad preexistente en algún reino de verdades hasta entonces inaccesible. Pensar una cosa u otra es relevante para la matemática que se desarrolle, pero los mismos desarrollos no lograrán mostrar como verdadero o falso a ninguno de los compromisos en los que se basa.

Puede decirse, por tanto, que cabe una atribución directa y una atribución indirecta de los aspectos conscientes y de los aspectos inconscientes

en la metáfora homo-máquina. Los aspectos conscientes son directamente atribuíbles al homo e indirectamente atribuíbles a las máquinas. P. ej., mientras se dice que un homo razona, de una maquina se dice que simula tal o cual razonamiento. Los aspectos inconscientes, sin embargo, sólo son atribuíbles al homo. En efecto, a un humano puede atribuirsele una líbido exagerada, pero esta cualidad no la puede tener, y difícilmente la puede simular, una máquina.

La distinción de lo mental en raciónal e irracional debe ser útil en la clarificación de algunas afirmaciones que se han producido en el ámbito de la IA. Una rama de la misma vista freciuentemente como competidora de facultades humanas es la de los SBC, en tanto simulan, con razonable éxito, destrezas inteligentes relacionadas con la capacidad raciocínica del homo. A partir de ahora, los comentarios sobre la IA se referirán especialmente --y aún sin indicarlo-- a esta rama suya.

4. Objetivo

El objetivo de este trabajo es tratar de centrar algunas afirmaciones extremas en torno a la IA y, en concreto, sobre aquellos ámbitos suyos de desarrollo que, como los SBC, desmitifican alguna parcela narcisistamente humana, si el homo considera a la inteligencia y a una de sus manifestaciones, la capacidad argumentativa, como una característica propia e inimitablemente suya. Este estudio es necesario, porque tanto filósofos (J. Searle), como ingenieros del conocimiento (M. Minsky) han vertido juicios exagerados acerca de las limitaciones o posibilidades de la IA. Sus mensajes parecen destinados, más que a clarificar algún aspecto de la disciplina, a demarcar la diferenciabilidad de paradigmas cognitivos. En filosofía, la confianza de que existe un núcleo de problemas inabordable por métodos físicos u objetivales; en ingeniería, la impresión de que la IA progresa a un ritmo que hace augurar que todos los fenómenos hoy considerados subjetivos, serán simulables en un futuro cercano.

5. Un marco de discusión

Para analizar la repercusión que tienen algunas afirmaciones hechas en el ámbito de la IA sobre distintos aspectos de la inteligencia humana (en tanto que la primera simule correctamente aspectos de la segunda) es

conveniente situar la discusión en un marco apropiado. Antes de delimitar este marco, algunas observaciones de tipo general pueden ser hechas:

(a) La comparación IA-inteligencia humana no es simétrica.

En efecto, lo que se compara son los productos de IA con lo que consideramos actitudes inteligentes del ser humano, no al revés, ya que la inteligencia humana constituye el espejo en el que se miran estas máquinas. Hay excepciones a esta regla, aunque son restringidas: el caso más común es cuando se comparan las destrezas humanas para resolver operaciones muy complejas, aunque perfectamente definidas (algorítmicas) (p. ej., una raíz cuadrada suficientemente complicada), con la que tiene un ordenador. En este caso, es obvio que una simple calculadora toma ventaja. Pero en IA lo importante no es tanto la capacidad de resolver problemas algorítmicos, cuando ofrecer soluciones a problemas que tiene un gran número de reglas implícitas, algunas de ellas bastante imprecisas.

(b) Importa entender qué se considera por 'inteligencia artificial', en el bautismo no demasiado afortunado de esta disciplina.

La definición de inteligencia artificial es importante, pues debe ayudar a limitar su ámbito de estudio, así como a señalar los límites que otras ciencias tienen en su área temática. Caben distintas definiciones de IA, pero todas ellas parecen deudoras de aquella que, a mediados de siglo, propuso Turing. En su conocido trabajo, *Computing machinery and Intelligence*, establece el principio de que una máquina es inteligente si supera su test, el test de Turing. Una máquina supera el test cuando, en su tarea de sustituir a un humano que pretende engañar a un entrevistador sobre su sexo, el porcentaje de errores que comete el entrevistador es similar cuando actúa el humano o su sustituto, la máquina.

Interesa llamar la atención sobre lo siguiente: en el test de Turing, la máquina no simula nada que no sea proferencia hablada o texto escrito de un proceso mental: lo que el individuo cree que debe decirle al interlocutor para engañarle. Lo que la máquina reproduce es una conducta lingüística, no las intenciones del sujeto, aunque evidentemente, esas intenciones puedan observarse, si el programador ha sido suficientemente hábil, en los comunicados de la máquina (p. ej., si ésta, simulando a una mujer, escribe: 'sabrías que soy hombre si pudieses tocar mi biceps'). Esta frase traduce una intención: la de mentir. Pero obviamente la máquina no

miente. Simula una frase cuyo significado es falso, no los factores mentales que fueron origen de esa frase.

(c) Importa caracterizar, aunque hacerlo de modo completo es imposible, qué se considera por inteligencia humana y, desde el punto de vista de la IA, en función de qué se escogen algunos atributos como constituyentes suyos.

Dar una definición de inteligencia humana es complicado. Aquí esta tarea es descargada en parte porque tal definición atiende a una perspectiva determinada: la de la IA. Desde este punto de vista, algunas características que normalmente se asocian a la inteligencia humana son las siguientes:

(1) Encontrar soluciones a problemas que no están totalmente definidos.

(2) Discriminar aspectos similares en objetos diferentes.

(3) Discriminar aspectos diferentes en objetos similares.

(4) Saber razonar con mensajes contradictorios o ambiguos.

Si la inteligencia humana se observa así, su simulación lo será, en buena medida, de estas características, notoriamente vinculadas al carácter argumentativo del lenguaje humano o a alguno de sus beneficios, como la capacidad discriminatoria del entorno en el que nos movemos. Por tanto, si se dice que la IA simula la inteligencia humana hay que tener en cuenta que no se intenta reproducir cualquier aspecto de esta inteligencia, sino sólo aquellos contenidos que el lenguaje transmite con suficiente estructuración racional. Según esto, no tiene sentido plantear, p. ej., si una computadora puede simular la actividad onírica de un sujeto (salvo que esta actividad sea reconstruída racionalmente). En este contexto, también es una acusación frecuente la de que hay aspectos de nuestro comportamiento racional que no simulan las máquinas de la IA. Podría decirse, p. ej., que la defensa de la reconversión industrial que el Sr. Ministro de Industria hace ante un grupo de mineros comporta prácticas inteligentes que una máquina nunca realizará, pero no por ello se puede decir que las máquinas no pueden ser ministros. Son otras las causas. Puede ser discutido si una máquina puede hacer esta tarea --seguramente, no habría grandes objeciones en contrario, pero esta discusión, más allá de la curiosidad intelectual, sería yerma, ya que en el contexto de nuestra cultura éste no es un aspecto interesante a reproducir por máquina alguna.

La IA simula preferentemente aquello que puede ser socialmente útil. Estas consideraciones deben llevarnos a matizar dos formas posibles de entender esta disciplina.

6. Otra distinción útil en IA

Como es sabido, 'Inteligencia Artificial' es un término acuñado por McCarthy en la década de los cincuenta para inaugurar un nuevo saber: aquel que debería ocuparse de estudiar las afinidades entre el cerebro humano y la máquina electrónica. El bautismo no fue demasiado afortunado, ya que el uso del término 'inteligencia' provocó un sarpullido que se reflejó --y sigue todavía haciéndolo-- en afirmaciones un tanto extremas, consecuencia, quizás, de las inespecifidad de esta palabra y de su papel en la definición del homo como especie-específica. Recientemente, y debido al desarrollo de los SBC o Sistemas Expertos (SE), se tiende a sustituir el término 'Inteligencia Artificial' por el de 'Expertos Artificiales', de menor escozor que el anterior, si bien más restringido, ya que no se incluyen en este campo partes de la IA que han tenido un notable desarrollo y que no tienen la estructura los SBC. Más adelante discutiremos algunos aspectos relativos a este tema a la luz de la distinción que se introduce a continuación, que deberá sustituir a la dicotomía IA dura/IA blanda, de escasa adecuación explicativa para las afirmaciones en torno al proyecto IA.

Bajo el nombre genérico de IA, se pueden distinguir dos grandes áreas:

(1) Inteligencia artificial pura, cuyo objeto de estudio se centra en debatir si se pueden conseguir computadores que tengan una inteligencia similar a la humana. Los temas de discusión en este ámbito girarán en torno a la cuestión ¿es ésta empresa posible? Preguntas y respuestas dentro de este apartado son de índole teórica.

(2) Inteligencia artificial aplicada, que tiene por objeto construir artefactos que de hecho resuelvan problemas para los que se suponen es requerida alguna destreza de la inteligencia humana. Lo que en este ámbito se plantee debe responder a la cuestión ¿es este intento útil? Preguntas y respuestas dentro de este apartado pueden ser de índole teórica o aplicada.

Esta distinción, como hemos anunciado, debe mostrarse pertinente para encuadrar correctamente algunas afirmaciones acerca del proyecto IA,

ayudando a dirimir cuál puede ser su alcance. Toda cuestión acerca de si es posible construir una máquina que simule, en general, la inteligencia humana debe situarse en (1). Las cuestiones acerca de si hay o es posible construir a corto plazo máquinas que simulen de hecho alguna capacidad humana se encuadran en (2).

La diferencia entre IA pura e IA aplicada marca también, de alguna forma, lo que ha sido el inicio y el desarrollo de la IA, la prehistoria y la historia de este proyecto. La prehistoria viene señalada por una idealización del proyecto, pudiendo servir de ejemplo el optimismo de Hilbert sobre la posibilidad de que cualquier discurso racional coherente pudiese ser axiomatizado con herramientas lógicas. La solución a cualquier problema se plantearía, entonces, como la necesidad de encontrar, una vez fijados los axiomas, la regla que permitiría llegar al objetivo. Los teoremas de Church y Gödel echaron por tierra esta idea; arruinaron la convicción de que todo problema planteado en un discurso estructurado tenía una solución general y algorítmica.

A partir de entonces la IA se centra en el desarrollo de soluciones a problemas concretos; entra propiamente en su historia. Los esfuerzos se encaminan, no en la dirección de subsanar las deficiencias anotadas a los resolvedores generales de problemas, sino en disponer de ingenios que solucionasen cuestiones concretas. El intento es sin duda más modesto, pero a la larga se mostrará más fructífero, al promover una nueva concepción del software: se intentan conseguir formas de representación del conocimineto no algorítmicas y diseñar programas que actuen deduciendo y razonando más bien que computando. La consecuencia fue que, en vez de demostradores generales de teoremas, se lograron SBC, dando contenido a la quinta generación de ordenadores.

La diferencia entre prehistoria e historia de la IA apunta también a distintos tipos de preguntas que se pueden hacer sobre su ámbito de estudio. A continuación ponemos tres tipos de cuestiones paradigmáticas, muy genéricas, que se han hecho o están implícitas en bastantes trabajos o afirmaciones acerca del proyecto IA. La categorización pura/aplicada debe mostrar tanto el tipo de respuesta que debemos esperar de estas preguntas como el alcance de sus predicciones:

1. *¿Pueden pensar las máquinas?* Pregunta en el ámbito de la IA pura. Debe esperarse una respuesta en este mismo ámbito.

2. *¿Puede construirse una máquina que simule tal o cual comportamiento del experto humano?* Pregunta en el ámbito teórico de la IA aplicada. Debe esperarse respuesta en el ámbito puro o aplicado de la misma.

3. *¿Hay, o está a punto de diseñarse, una máquina que simula tal o cual comportamiento inteligente humano?* Pregunta en el ámbito de la IA aplicada; respuesta en este mismo ámbito.

Pues bien, esta distinción sobre posibles preguntas acerca de la IA (y,por tanto, con su posible metaforización, acerca de funciones cognitivas humanas) debe, a nuestro juicio, ofrecer un marco de discusión adecuado para situar correctamente el alcance de afirmaciones o predicciones sobre este tema. El análisis de algunos ejemplos deberá insinuar los beneficios aquí prometidos.

Pero antes detengámonos en la pregunta-tipo número 1, muy relevante desde el punto de vista de la IA pura y de la lógica, uno de los saberes colaboradores. Según la categorización introducida, la cuestión se plantea en el terreno de la IA pura y es ahí donde se debería esperar una respuesta. No obstante, en círculos de divulgación científica se ha tendido a trasvasar las consecuencias que se siguen de algunas significadas respuestas al ámbito de la IA aplicada.

Como es conocido, la pregunta ¿puede pensar una máquina? fue paradigmáticamente planteada por A.M. Turing en 1950. Ya entonces señala el autor que la objeción más seria a una posible respuesta positiva la constituían los teoremas limitativos de Gödel. Analicemos pues, brevemente, la objeción de Gödel.

7. Objeciones a la posibilidad de máquinas totalmente inteligentes

Cuando se plantea la pregunta de si las máquinas pueden pensar o, más adecuadamente, si las máquinas pueden simular de modo completo la inteligencia humana, si no se limita el alcance del concepto 'inteligencia', debe entenderse que se está planteando si es posible o no una máquina que simule toda la inteligencia humana. Así entendida, esta pregunta puede tener dos respuestas en ámbitos de conocimientos distintos:

(a) En el ámbito de la ingeniería genética. Habría que responder a la cuestión de si es posible hacer clonos de seres humanos con sus mismas características.

(b) En el ámbito de la IA pura. Hay que responder a la pregunta de si es posible construir una biomáquina. Esta última cuestión involucra a su vez a otras dos: (i) ¿Puede realizar la máquina todas las funciones de un homo, incluídas las inteligentes?, (ii) ¿Tiene las propiedades de un homo que tienen que ver, como sustrato, con su conducta inteligente --el cuerpo, p. ej.?

La cuestión de si una máquina puede realizar todas las funciones inteligentes de un ser humano ha recibido, al menos, dos tipos de respuestas, la cásica de los teoremas limitativos de Church y Gödel y la más reciente de R. Penrose.

(1) La demostración de Churh y Gödel de que existe una proposición que puede ser 'vista' como verdadera por un homo pero que no puede ser demostrada como tal es una respuesta negativa a la cuestión que ocupa este apartado. En este caso, la negativa lo sería respecto al software que se pudiese construir: sería imposible confeccionar un programa que acumulase todas las proposiciones verdaderas, cúmulo que, desde un punto de vista teórico, sí puede hacer un humano.

(2) En el trabajo de R. Penrose, la limitación a la que se alude no sería respecto al software, a los programas, sino al hardware: el argumento, grosso modo, se basa en la consideración de que mientras el funcionamiento electroquímico del cerebro se rige por leyes de la física cuántica, donde la indeterminación tiene un gran papel, los circuitos de las máquinas operan según una física de estados discretos.

Tanto la objeción de Church-Gödel como la de Penrose se sitúan en el ámbito de la IA teórica y las limitaciones que apuntan deberían restringir su alcance a este mismo ámbito. Veamos esto con un poco más de detenimiento.

A. Objeción a la posibilidad de conseguir un software que simule toda la inteligencia humana. La objeción de Church-Gödel.

Esta objeción tiene un planteamiento adecuado en el área de la IA conocida como demostración automática de teoremas (DAT). De un modo general, se puede decir que el objeto de la DAT es estudiar métodos al-

gorítmicos para decidir si una fórmula F es derivable de un lenguaje formal LF. Como es conocido, en lógica de primer orden, existen métodos para decidir si F es o no un teorema de LF, pero no existe un método general que resuelva el problema en todos los casos. Veámoslo.

La demostración de Gödel se basa en el proceso conocido como aritmetización de la lógica, mediante el cual es posible representar a través de un número (número de Gödel) símbolos y fórmulas de un lenguaje formal y, también, pruebas efectuadas con este lenguaje. F sería derivable de LF si el número Gödel que representa a F fuese un número perteneciente al conjunto de números Gödel de los teoremas de LF. El conjunto de números Gödel de los teoremas de LF tiene la propiedad de ser recursivo, esto es, dado un número tenemos un método efectivo --una máquina de Turing, p. ej.-- que nos dice si es o no un número perteneciente a uno de esos conjuntos de números.

Podría pensarse, entonces, que una de las consecuencias de la gödelización es la posibilidad de tener una solución universal para demostrar teoremas. Dada una fórmula, determinaríamos su número Gödel, lo introduciríamos en una máquina y ésta nos diría si pertenece o no al conjunto de números de teoremas. Pero esta previsión falla porque ni el conjunto de los números que representan a los axiomas ni el de los teoremas son siempre recursivos.

Por tanto, incluso en un ámbito tan bien delimitado y preciso como el de la aritmética hay limitaciones en la decibilidad. Estas han quedado recogidas, de un modo exhaustivo, en los siguientes teoremas:

1. Teorema de Church. Si E es una extensión consistente de la aritmética, entonces E no es decidible.

2. Primer teorema de incompletud de Gödel. Si E es una extensión axiomatizable de la aritmética, entonces no es completa; esto es, hay una fórmula tal que ni ella ni su negación son teoremas de E.

3. Segundo teorema de incompletud de Gödel. Si la aritmética es consistente, entonces no puede probarse dentro de ella la fórmula $\exists$ x (F(x) $\wedge$ $\forall$ y (- P(y,x)) donde P(y,x) es el predicado aritmético "y es el número de Gödel de una prueba de la fórmula que tiene como número de Gödel a x", y F (x) es "x es el número de Gödel de una fórmula".

El examen de estos resultados bajo la tesis de Church, que dice que 'calculable implica computable' o, por contraposición, que 'no computable implica no calculable', nos permite el paso inverso al de la gödelización; esto es, hablar, en función de los resultados obtenidos con los números, otra vez de fórmulas. De este modo se puede mantener que los teoremas de Church-Gödel, en el ámbito de la DAT muestran que,

1* La DAT no posee un método general efectivo para decidir si una fórmula es consecuencia de sus propios supuestos o axiomas.

2* Hay al menos una proposición en el lenguaje de DAT que, de acuerdo con los criterios de la inteligencia humana es verdadera, pero DAT no puede probar; por tanto, por la tesis de Church, que no puede decidir si es verdadera o no

3* Si DAT es consistente, no tiene técnicas generales efectivas para probar que lo es.

Estas consecuencias deben conducir a expresar la imposibilidad teórica de construir, de modo absoluto, máquinas inteligentes. Es ésta la conclusión que debe sacarse y no otra. La pregunta había sido planteada en el terreno de la IA pura y es respondida en este mismo ámbito.

Pero en ocasiones, como indicábamos anteriormente, se ha tendido a trasvasar el alcance de estas conclusiones al ámbito de la IA aplicada ¿Influyeron realmente los resultados de Gödel en la evolución de la IA de un área (los DAT), a otra (los SBC)? Esto parece cuestionable, ya que los teoremas de Gödel no demuestran que no se pueda probar como verdadera toda proposición que lo es, sino que no se pueden probar como verdaderas, a la vez, todas las proposiciones que lo son. La máquina puede simular cualquier tipo de comportamiento inteligente, pero no todo el comportamiento inteligente. De ser viable la construcción de un resolvedor general de problemas, este proyecto no se habría paralizado por los teoremas limitativos, ya que sería enormemente útil tener un artilugio tal, aunque no fuese capaz de resolver todos los problemas a la vez. El abandono de esta idea se debió, más bien, a otras dificultades (p. ej., la explosión combinatoria), no a los problemas teóricos señalados.

Esta idea se puede afianzar, además, en los propios resultados de Gödel. Se puede observar que la proposición indemostrable $\exists$ x (F(x) $\wedge$ $\forall$ y (- P(y,x)) no tiene nada que ver con el enunciado de los teoremas de la

aritmética, como el teorema de descomposición en fáctores primos utilizado en la gödelización. Para las partes útiles de la aritmética se han construído cálculos completos en los que es posible derivar todos los teoremas interesantes.

B. Objeción a la posibilidad de conseguir un *hasdware*, un dispositivo electrónico, que simule al cerebro humano.

Esta tesis ha sido defendida recientemente por R. Penrose en su libro *La nueva mente del emperador*, que será comentado muy brevemente. El interés de hacerlo se justifica en las espectativas que, de nuevo, se han generado en algunos ámbitos acerca de su posible repercusión en el desarrollo del proyecto de la IA aplicada. De un modo breve, su tesis puede resumirse así: la física del cerebro es de naturaleza cuántica. Su funcionamiento, por tanto, es no determinista. En cambio, la física de un ordenador no es cuántica y sus patrones de funcionamiento son algorítmicos. Ahora bien, como muestra el teorema de Gödel, existe un ingrediente noalgorítmico esencial en los procesos de pensamiento consciente. Ergo, un ingenio electrónico no puede ser, por definición, igual a un cerebro humano.

Después de una crítica a la interpretación no realista de la mecánica cuántica, Penrose plantea la posibilidad de que, tal vez, en la investigación en IA habría que tratar de construir computadores con soporte físico cuántico. No obstante, en su opinión, debería darse todavía algún cambio sustancial en la teoría cuántica clásica para que se pudiese, siquiera balbucear la idea, de explicar con ella la naturaleza de la mente.

Las afirmaciones de Penrose deberían encuadrarse en el apartado que hemos denominado como 'IA teórica'. Su relevancia desde este punto de vista, no obstante, está todavía por ver, ya que el estudio de Penrose adolece del carácter conclusivo y de los comentarios extensivos de los que sí ha sido objeto el trabajo de Gödel (que él, por otra parte, utiliza de forma nuclear en la articulación de sus tesis).

En cualquier caso, si la tesis de Penrose fuese correcta, tendría una repercusión importante en la IA teórica, pero mostraría escasa importancia para la IA aplicada. Todos sabemos que el avión simula bien la cualidad voladora del pájaro o el tractor la cualidad roturadora del buey. Pero con la misma evidente facilidad podemos constatar que la físico-química del organismo del pájaro no es la misma que la del avión y que la del buey es

diferente de la del tractor. Para hacer una buena simulación, el tractor no tiene por qué tener la anatomía ni la fisiología de un buey, ni el avión el cuerpo de un pájaro. El ejemplo de Gunderson (Gunderson, 1964) acerca de cómo una bolsa de piedras podría jugar al juego de la imitación, es una réplica interesante, desde la IA teórica, a los argumentos que discuten si, para simular una conducta inteligente, se requiere un soporte similar al humano. En un ente simulador, el soporte puede ser otro.

En IA aplicada hay un caso que ejemplificaría paradigmáticamente cómo puede darse una simulación relativamente satisfactoria de un fenómeno, sin que ello arroje luz acerca de la naturaleza de ese fenómeno. Hablamos de las redes neuronales, programas informáticos que intentan simular las conexiones cerebrales humanas que producen actividades inteligentes. Es sabido que las redes aprenden con facilidad algunas tareas a partir de otras que han sido especificadas previamente. Pero, cómo aprenden las neuronas informáticas es en buena medida una caja negra todavía por descifrar. En este caso, pues, se da una relativa simulación del aprendizaje humano, pero la simulación no explica nada del mismo, ya que se desconoce cómo se produce la red.

Las consideraciones hechas hasta aquí deben ser suficientes para abordar el último apartado: el análisis de dos ejemplos, que responden a alguna de las preguntas-modelo planteadas antes.

8. Análisis de dos ejemplos

En este apartado serán comentadas algunas frases-tipo sobre la posible inteligencia de las máquinas. Se han escogido textos extremos, que pretenden ser representativos del enconamiento que a menudo se ha producido y todavía se produce en estas discusiones. La distinción entre IA pura/IA aplicada debe mostrar ahora algún beneficio, categorizando las afirmaciones y mostrando su importancia y alcance dentro de cada categoría.

Se han escogido comentarios de un filósofo (J. Searle) y de un ingeniero informático (M. Minsky). Ambos tienen un carácter predictivo: en el caso del filósofo, la predicción es negativa; al modo wittgensteniano, señala lo que no puede ser. La del ingeniero es una predicción positiva, ofrece un vaticinio de cuándo será.

A. La frase de J. Searle es bastante conocida, debido al éxito que ha tenido el libro donde aparece, así como al carácter divulgativo de las "*Reith Lectures*", foro en el cual fue originalmente proferida. Dice:

> En una palabra, la mente tiene más que una sintaxis, tiene semántica. La razón por la que el programa de computador no puede jamás ser una mente es simplement 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 (Searle, 1984: 37)

Conforme a la categorización establecida, la afirmación de Searle se encuadra dentro del apartado IA pura. Sus repercusiones, de tenerlas, lo serían en este ámbito y, en principio, no deberían trasvasarse consecuencias a la IA aplicada. Sería posible hacerlo bajo la presuposición de que los científicos de la IA aplicada tienen como objetivo construir un ordenador humano. Pero, como ya se ha indicado, esta suposición es falsa; no es programáticamente correcto, ni entra dentro de los planes de la IA aplicada, construir máquinas humanas, sino diseñar máquinas útiles en tanto que útiles son algunos de nuestros conocimientos.

La tesis de Searle encuentra justificación es su experimento teórico concido como 'habitación china', del que damos una breve noticia a continuación: supongamos que una persona responde a las preguntas en lengua china de un entrevistador, profiriendo casualmente frases que ha aprendido simplemente a repetir, pero cuyo significado desconoce. Sin embargo, sus frases coinciden exactamente con lo que el entrevistador esperaba que le fuese contestado. El entrevistador puede pensar legítimamente que ésta persona sabe chino, pero un observador externo no estaría dispuesto a decir que esa persona comprende esta lengua. La moraleja de este relato es la siguiente: una máquina puede ejecutar adecuadamente un algoritmo sin que ello quiera decir que tenga comprensión de él. Esto le ha servido a Searle para señalar que los computadores no podrán ser nunca inteligentes. Pues bien, aunque esto es sin duda interesante para debatir desde un punto de vista teórico, desde un punto de vista aplicado la observación de Searle carece de interés ya que, como se ha indicado en la definición programática que de la IA hizo Turing, la tarea asignada a las máquinas era la de simular, no la de comprender. Cualquier cuestión

acerca de si la simulación en una máquina de un comportamiento comprendido por algún humano implica algún tipo de comprensión por parte de la máquina misma, es un tema interesante para el debate teórico, pero viciado por las creencias subjetivas de quien responda a esa cuestión. Desde un punto de vista aplicado, como ya se indicó anteriormente, constituye un problema que, de momento, no tien respuesta.

Respecto a su afirmación de que 'los ordenadores no tienen semántica' puede decirse lo mismo: la frase es relevante teóricamente, pero no supone ninguna limitación práctica para los ordenadores. En efecto, supongamos que tenemos una regla de Modus Ponens en un motor de inferencia y en la base de datos un aserto antecedente que hemos catalogado como bastante cierto. La regla, si está bien caracterizada, nos permite deducir un consecuente que será valorado, por lo menos, como tan 'bastante cierto' como el antecedente lo es. Si convenimos --de un modo muy simplificador respecto a las cualidades semantizadoras de un ser humano-- es decir que el significado de una proposición es algún valor de verdad, las reglas de un SE no sólo deducirían sintácticamente nuevas entidades formales, sino que transmitirían, todo lo simples que se quieran, significados (vagos). Evidentemente el ordenador no tiene semántica propia, pero sí la que el programador ha sabido darle (en este caso, una semántica vaga).

Por tanto, si bien las afirmaciones de Searle son notables para la discusión de problemas en el ámbito especulativo de la IA, no han tenido ninguna influencia en la IA aplicada.

B. Analicemos ahora algunas frases de M. Minsky, uno de los personajes más conocedores, influyentes y polémicos de la llamada IA dura. Como es conocido, Minsky ha hecho importantes aportaciones en este campo, tanto en el tema de la representación del conocimiento, como en el de la simulación neuronal. Como reacción al libro de R. Penrose, ha manifestado recientemente en una entrevista periodística --lo que puede explicar la dureza de sus afirmaciones-- que

1. El cerebro es una máquina terriblemente complicada, como mil ordenadores todos distintos; de momento sólo entendemos unas pocas funciones, pero en poco tiempo, quizá 100 años, es posible que las entendamos todas (Minsky, 1991).

2. La religión promete una vida futura, pero la IA supone que algún día podrás continuar tu vida presente. Se conservará en disquete la personalidad de los que mueran: eso equivale a la inmortalidad, y con el añadido de que se podrán hacer tantas copias como se quiera (Minsky, 1991).

Comentémoslas separadamente.

La primera frase tiene como ámbito de discurso a la IA aplicada y , sin embargo, nos es estrictamente hablando una afirmación razonable apara este ámbito. Es una predicción que ya fue hecha en otras ocasiones más ingenuas y eufóricas de la IA, de la cual sólo puede decirse que es muy aventurada. Como es conocido, una metáfora repetida en los inicios del proyecto IA se refería a que el cerebro humano era tan complejo que su simulación sólo podría venir dada por un cúmulo de ordenadores que ocupase la extensión de un campo de fútbol. Pronto se observó que el problema mayor no era el de disponer de muchos ordenadores convenientemente conectados, sino la especialización de los mismos. La dificultad para simular en el ordenador, no las facultades superiores del homo, sino sus comportamientos más ordinarios e ingenuos, como p. ej., aprender a freir un filete de ternera después de saber freir un filete de pollo, hace que se vea como muy improbable la realización de la predicción de Minsky. Es más: se podría decir que de realización imposible, aunque esto también es una conjetura arriesgada. La razón que aducimos es que, de conseguir simular todas las funciones cerebrales humanas, habríamos simulado todas las funciones concocidas, pero, como afirman los neurofisiólogos, todavía queda mucho por conocer del cerebro. La afirmación de Minsky presupone que el programa de la neurobiología siga un desarrollo lineal y esté acabado dentro de 100 años, tesis que posiblemente no haría suya ni el más optimista de los neurofisiólogos.

Su segunda afirmación cae dentro del terreno de la IA pura. Es una frase también predictiva, pero a diferencia de la anterior, debe señalarse su carácter religioso. Su afirmación tiene la misma impronta que las frases de cualquier texto revelado. En efecto, en ella no sólo se aventura que se conseguirá un logro científico (las herramientas necesarias para copiar la personalidad humana), sino también, un éxito metafísico (su copia y desdoblamiento efectivo). Si ello fuese posible, tendríamos máquinas con personalidad melancólica, autoritaria, irascible, etc. Pero si no se especi-

fica lo que se entiende por 'personalidad' y se usa a esta palabra en su sentido normal, deberíamos convenir en decir que la oración 'Esta máquina tiene personalidad' es , a la luz de nuestro conocimiento, una oración anómala, como anómala es la frase 'Tengo recuerdos azules'. Como ya indicamos antes, se puede, en cierto sentido, asociar a un robot un programa de personalidad neurótica, que ocasione un comportamiento mecánico que pueda ser catalogado como tal. Pero el robot no tendría personalidad neurótica, sólo podría simular lo que por comportamiento neurótico entiende el especialista, conocimiento que el informático habría implementado en su programa.

Para la inmortalidad que promete Minsky no se necesita desarrollar el programa de la IA. Se ha dicho en numerosas ocasiones que alguien que escribe sus ideas se inmortaliza en el papel. Con los modernos métodos de reprografía, se pueden hacer tantas copias de esa personalidad como se quiera. Pero obviamente, todos consideramos que este es un uso metafórico del concepto de inmortalidad. Su frase, por tanto, sólo es relativamente relevante para el debate teórico y escasamente informativa para la IA aplicada, por la excesiva audacia de la conjetura que propone.

9. Conclusión

Este trabajo ha tenido como objetivo proponer un marco en el que clasificar algunas afirmaciones sobre la IA modificando uno preexistente: aquel que las separaba en pertenecientes a la IA dura o a la IA blanda. Creemos que la división entre IA pura e IA aplicada es más útil para entender, en su justo sentido, algunas frases que, acerca de este tema, se han producido en diversos ámbitos de conocimiento. Aquí se han escogido citas de J. Searle y M. Minsky, y las hemos analizado con la convicción de que constituyen buenos elementos representativos de pléyades de frases que comunican mensajes análogos: la imposibilidad de simular el comportamiento humano o la confianza certera de que esto se hará. Hemos mostrado que las afirmaciones de Searle y Minsky son comentarios exagerados que, de no ser contextualizados en alguno de los ámbitos establecidos, pueden ser considerados simplemente como falsos. Por tanto, la categorización introducida ha debido aportar elementos estructurales interesantes para dimensionar, en un sentido realista, lo que cada frase puede transmitir.

Esta misma clasificación debe hacer ver cómo es posible proferir frases adecuadas en cada uno de los apartados considerados.

(i) Como frase paradigmáticamente interesante para el apartado de IA pura proponemos la que enuncia el teorema de Gödel, que, efectivamente, en el ámbito de discurso de los lenguajes formalizados, señala una limitación teórica a la IA. Es una frase que tiene un contexto específico y un alcance concreto.

(ii) La definición de Duda y Shortliffe de SBC como

> un programa de IA cuya acción depende más de la presencia explícita de un amplio cuerpo de conocimiento que de la posesión de procedimientos computacionales ingeniosos (Duda & Shortliffe, 1983: 267)

puede servir como paradigma de afirmación teórica intereesante en el ámbito de la IA aplicada.

(iii) Por último, el siguiente comentario de Tennant puede servir como muestra de discurso interesante para la IA aplicada:

> Los niños de tres años entienden el lenguaje bastante bien y según crecen su lenguaje va mejorando. En una época posterior es cuando pueden aprender el cálculo matemático, a diagnosticar enfermedades y cosas por el estilo. Sin embargo, es mucho más fácil construir un sistema que pueda hacer cálculo matemático que uno que entienda el lenguaje. La razón estriba en que el entendimiento del lenguaje natural requiere una enorme cantidad de conocimiento: conocimiento acerca del lenguaje y de la comunicación en general, pero fundamentalmente conocimiento del mundo. Debido a esta cantidad de conocimiento necesaria para comunicarse a través del lenguaje, mi expectativa para una comunicación tipo Star Trek es que será uno de los últimos problemas resueltos por la IA. La pregunta fundamental es si va a ser económicamente viable el poner esa cantidad inmensa de conocimiento dentro de un ordenador. Estimo que nadie querrá pagar por codificar todo el conocimiento que posee un niño de tres años que, por cierto, es una cantidad increíble... Es económicamente más rentable el implementar el conocimiento de un experto que el de un niño de tres años; por tanto, actualmente no hay una motivación económica para realizar esta tarea (Tennant, entrevista en Mishkoff, 1985: 99).

En resumen, creemos que en la discusiones acerca de la IA se debe evitar la devaluación o transgresión semántica de palabras como 'inteligencia', 'pensar' o 'personalidad', algunas de las cuales hemos analizado en este texto. El uso habitual y normalmente descuidado de las mismas en multitud de trabajos sólo se entiende desde un punto de vista metafórico. La metáfora quizás sea un recurso literario adecuado para divulgar el proyecto y los objetivos de la IA, pero su uso debe evitarse, en la medida de lo posible, en contextos científicos. La asociación ilegítima de palabras puede llevar a la ilusión de que una entidad se predica de un término y que esa predicación está respaldada por la investigación corriente. En contextos científicos hay sustitutos para palabras como las de arriba que tienen menos cargas ontológicas y que, además, ayudan a entender mejor de qué se está hablando. P. ej., en vez de 'inteligencia', 'experiencia', en vez de 'pensar', 'simular', en vez de 'personalidad', 'software que personaliza', etc. Todas ellas, no obstante, son susceptibles de clarificaciones teóricas, que en este ámbito se hacen necesarias y útiles.

Palabras como 'inteligencia' o 'pensamiento' son frecuentemente asociadas --y también lo han estado es este trabajo-- a los aspectos racionales del homo. Pero otro componenete sustancial de la mente humana son los aspectos irracionales ¿Qué podemos decir de los mismos en la etapa final de este trabajo?

El tema del insconsciente --recordamos-- nos vincula a la ofensa psicológica tal y como la relató Freud. Pero después de haber analizado la otra posible ofensa proviniente también de la psicología --de la cognitiva, podemos decir que el motivo de la ofensa en el paradigma freudiano --el inconsciente-- se puede convertir en paladín defensor de alguna parcela típicamente humana. La simulación de algunas facultades conscientes del homo ha hecho que la IA haya representado una afrenta a algunas propiedades tenidas como especie-específicas. En este sentido, algunas facultades razonadoras han sido reproducidas con resultados limitados aunque esperanzadores. En cambio, las facultades irracionales, aquellas que no están sometidas a reglas o lógica alguna, parecen ser un reducto seguro del homo en su intento por mantener una entidad diferenciada. Por tanto, la objeción de Freud al narcisismo humano se ha convertido, gracias a la IA, en su seña de identidad más segura.

No obstante, la simulación de lo racional no debe ser considerado como un ataque a una cualidad humana. No por ser simulada la cualidad deja de

ser humana. La simulación lo que pone de manifiesto es que, para algunas tareas (las que son objeto de mímesis) no se necesita el soporte humano; del mismo modo que para ir de un barrio de una ciudad a otro ya no son necesarias las piernas. Pero esto no supone una minusvaloración de la pierna como entidad corpórea; simplemente se indica su papel relativo para esta empresa.

La IA en general y los SBC en particular reflejan un sentir de nuestra época: el énfasis en el conocimiento y la valoración de la razón. En una reciente entrevista televisiva, A. Toffler señalaba, respecto a la evolución de la sociedad, que primero fue la violencia, después el dinero (etapa en la que todavía estamos) y que muy pronto vendrá la era del conocimiento. Apuesta porque en las sociedades modernas ésta sea la moneda de cambio más usada, resueltas algunas necesidades materiales con la ayuda de las máquinas. F. Sáez Vacas ha hecho especial énfasis también en este concepto, aunque en un contexto distinto, al indicar que en el desarrollo de la Informática primero fue la información, ahora es el momento del conocimiento y falta por alcanzar la etapa de la sabiduría (Sáez-Vacas, 1991). Esta tesis constituye, sin duda, una interesante predicción teórica en el ámbito de la IA aplicada.

REFERENCIAS

AMBLE, T. (1987): *Logic Programming and Knowledge Engineering*, Addison-Wesley, New York.

BOOLOS; G. S. & JEFFREY, R. (1974): *Computability and Logic*, Cambridge University Press, 1990[3].

CUENA, J. & CAMPBELL, J.A. (1989): *Perspectives in artificial Intelligence,* " Vols. Ellis Horwood, New York.

DUDA, R. O. & SHORTLIFE, E. H. (1983): "Expert Systems Research", *Science,* vol. 220, nº 4594: 261-268.

FREUD, S. (1917): "Una dificultad del psicoanálisis", en *Obras completas,* t. VII (1916-1924), Biblioteca Nueva, Madrid, 1974, pps. 2432-2437.

GUNDERSON, K. (1964): "El juego de imitación", en Anderson, A. R. (Ed.): *Controversias sobre mentes y máquinas*, Tusquets ed., Barcelona, 1984. Ver. original en *Mind*.

HAYES-ROTH, F. et al. (Eds.) (1983): *Building Expert Systems*, Addison-Weslwy, N.Y.

HÖFSTADTER, D. (1979): *Gödel, Escher y Bach. Un eterno y grácil bucle*, Barcelona, Tusquets Editores, 1987. Vers. original en Harvester Press, Hassoks, Sussex.

LEDESMA, L. de & LAITA, L. M. (1989): "Fundamentos lógicos de la Inteligencia Artificial", en *Curso de Conferencias sobre Inteligencia Artificial y Robótica*, Abril-Mayo de 1989. Real Academia de Ciencias Exactas, Físicas y Naturales, Madrid, pp. 70-84.

LUCAS, J. R. (1961): "Mentes, máquinas y Gödel", en Anderson, A. R. (Ed.): *Controversias sobre mentes y máquinas*, Tusquets ed., Barcelona, 1984. Ver. original en *Philosophy*, Vol. XXXVI.

MINSKY, M. (1985): "Nuestro futuro robotizado", en Minsky, M. et al.: *Robótica. La última frontera de la alta tecnología*, Planeta, 1986. Ver original en Omni Pub. International, New York.

MINSKY, M. (1991): "Entrevista", en *El País*, miércoles 4 de Diciembre de 1991, p. 28.

MISHKOFF, H.C. (1985): *A fondo: Inteligencia Artificial*, Ediciones Anaya Multimedia, 1988. Ver. original en Howard W. Sams & Co. Inc.

PENROSE, R. (1989): *La Nueva mente del emperador*, Biblioteca Mondadori, Madrid, 1991. Ver. original en Oxford University Press.

SAEZ-VACAS, F. (1991): "La sociedad informatizada. Apuntes para una patología de la técnica", *Claves de Razón Práctica*, Marzo, nº 10, pp. 74-94.

SEARLE, J. (1984): *Mentes, cerebros y ciencia*, Cátedra, Madrid, 1985. Ver. original en The 1984 Reith Lectures.

SEARLE, J. (1990): ¿Es la mente un programa informático?, *Investigación y Ciencia*, 162, pp. 10-16.

TRILLAS, E. (1989): "Una introducción breve a la lógica borrosa", en *Curso de Conferencias sobre Inteligencia Artificial y Robótica*, Abril- Mayo de 1989, Real Academia de Ciencias Exactas, Físicas y Naturales, Madrid, pp. 187-200.

TURING, A.M. (1950): "Maquinaria computadora e inteligencia", en Anderson, A.R. (ed.), *Controversias sobre mentes y máquinas*, Tusquets ed., Barcelona, 1984, pp. 11-50. Ver. original en *Mind*, Vol. LIX, nº 236.