

INTELIGENCIA ARTIFICIAL

Charla en el Ateneo de Madrid.

Calle Prado 21.

Día: 26 de octubre de 2012

Hora: 7:00 de la tarde

Lugar: salón CIUDAD DE ÚBEDA

Cuarta planta (hay ascensor)

Luis Ruiz de Gopegui

1. AGRADECIMIENTOS.

Muchas gracias por vuestra presencia en este acto. Gracias también al Ateneo de Madrid, a **Juan Fuertes** Director del la Sección de Ciencia y Tecnología y **Jaime Sánchez Montero** que al final conducirá el diálogo que abriremos si ustedes lo desean.

2. LOS COMIENZOS.

La gran aventura de las investigaciones sobre inteligencia artificial tomó forma, hace algo más de medio siglo, en una **Conferencia Internacional** que se celebró, en el verano de **1956**, en el Dartmouth College de Hanover, en **New Hampshire**, al Noreste de los EEUU. Antes había habido algunos escauceos por distintas partes del planeta debidos a investigadores e investigaciones independientes.

Entonces yo era un estudiante de la Universidad de Stanford, en California, en la otra punta de los EEUU. Cuando leí en las revistas científicas que circulaban por la biblioteca, el anuncio de aquella conferencia no me lo podía creer: al parecer unos cuantos científicos ilustres, mucha gente opinaba que estaban chiflados, se iban a reunir **“porque creían (en aquella época era sólo un acto de fe) que lo que llamamos pensamiento podía tener lugar fuera del cráneo humano, que podía ser entendido de una manera formal y**

científica, y que el mejor instrumento no humano para llevarlo a cabo era el ordenador digital”.

Los cuatro científicos que promovieron aquella conferencia eran, **John MCarthy**, un joven profesor de matemáticas de aquella universidad, **Marvin Minsky**, profesor de matemáticas y neurología en **Harvard**, **Nathaniel Rochester**, director de investigación en el Centro de **IBM** en Nueva York y **Claude Shannon**, matemático de los **Laboratorios Bell Telephone**, que ya era famoso por su **teoría estadística de la información**.

En la actualidad, debido a los grandes avances que se han logrado tanto en neurociencia como en ordenadores electrónicos, todo esto no llama mucho la atención ni **escandaliza**, sin embargo, en el año 1956, cuando la mayoría de ustedes no habían nacido, aquel inesperado inicio de la aventura de la inteligencia artificial, parecía una **tremenda provocación y hasta un acto blasfemo**, y digo blasfemo porque afectaba directamente a un tema tan delicado como el del **cuerpo y el espíritu**.

Recuerdo perfectamente como un compañero indonesio de la Universidad de Stanford y yo fuimos a ver a uno de nuestros profesores, se llamaba **Frederick Terman**, que nos daba clases de electrónica y que era extremadamente amable, más aún con los extranjeros, y le preguntamos si aquello tenía algún sentido. Él se sonrió y después de varios comentarios nos dijo en un tono de marcado escepticismo: “Nadie puede poner límites a la ciencia. A mí la idea de estos locos de Dartmouth me parece un poco **disparatada**, pero cuando yo era un chaval, en 1903, también me parecieron disparatados los experimentos de **Marconi** y fijaos hasta donde han llegado”.

Así es como empezó a consolidarse la inteligencia artificial, una ciencia que, creo que en aquella reunión, quedó **definida** como: **“La rama de la informática que estudia cómo programar a los ordenadores para que presenten un comportamiento en apariencia inteligente”**.

3. PRIMERAS REACCIONES.

Durante los primeros años de la larga travesía de la inteligencia artificial, el tema **pasó prácticamente inadvertido** para la mayoría de la gente, porque no se había enterado de tal acontecimiento o porque eran totalmente **escépticos** frente a sus propósitos. Sin embargo,

pasadas algunas décadas empezaron a aparecer en las revistas científicas bastantes artículos en los que se daba cuenta de algunos de los progresos que se estaban haciendo en ese campo. **Máquinas capaces de mantener una conversación** con cualquier persona, otras que traducían de un idioma a otro, etc..

Sobre estas máquinas que mantenían una conversación os quiero contar una anécdota. En un visita, de las muchas que he realizado a la **NASA** en Wasington, pregunté a un compañero de la Agencia sobre el tema de la inteligencia artificial y al día siguiente me facilitó una entrevista, con un experto en el tema. Éste científico, un chico joven, muy simpático y también muy bromista, me invitó a que mantuviera una conversación con un computador. Él se marchó del despacho y me dejó sentado delante de un teclado y una pantalla. En esa pantalla apareció una cara sonriente y debajo de ella una línea escrita en la cual me saludaba. Yo tenía que contestarle escribiendo con el teclado. Así empezamos una larga conversación, que duró algo más de media hora, transcurrida la cual el experto en inteligencia artificial volvió y me preguntó qué tal mi conversación. Luego estuvimos charlando sobre el tema y por fin nos despedimos. Cuando ya me marchaba, le pedí por favor si se podía borrar la conversación que había mantenido con la máquina. "Que más te da. Has estado hablando con un cerebro de hoja de lata, que no tiene sentimientos y no le importa nada lo que has dicho". "Sí estoy de acuerdo contigo, pero preferiría borrar la conversación". Se echo a reír y dándome una palmada en la espalda susurró. "**La máquina te ha hecho creer que era una persona**". En efecto, la máquina me fue enredando en la conversación y me obligó a contarle cosas que podían ser consideradas como "intimas".

Volviendo a los comienzos de la inteligencia artificial, a la vista de sus primero éxitos, enseguida empezaron a surgir **comentarios para todos los gustos**.

Los había totalmente **despectivos y escépticos**, principalmente por parte de algunos psicólogos y filósofos, un grupo de estudiosos que siempre han mostrado cierto recelo acerca de los avances científicos, porque en más de una ocasión les han dado algún revolcón. Que tontería - decían - una máquina de latón **nunca podrá enamorarse**, un ordenador electrónico nunca podrá **creer en Dios**, las máquinas de hoja de lata jamás podrán enunciar **conclusiones éticas**. Y otras cosas por el estilo. Pero sus críticas no tenían mucho sentido. **Los tiros no iban por donde ellos creían**.

Sin embargo, también los había del signo contrario y enormemente **alarmistas**, como el de un tal **Edward Fredkin**, profesor de ingeniería electrónica en el **MIT**, que dijo: **“Al final, hagamos lo que hagamos, habrán máquinas inteligentes con metas propias. Estoy plenamente convencido de ello. Es posible que se pueda retrasar. Tal vez hasta pueda evitarse, no lo sé. Pero es muy difícil que una máquina, un millón de veces más inteligente que tú, sea tu esclava”**.

4. LOS PRIMEROS PASOS.

Los comienzos de la inteligencia artificial fueron muy **tortuosos**, entre otras razones, porque se habían puesto demasiadas expectativas en esta nueva ciencia. Las **películas de ciencia ficción** habían pintado paisajes tremendistas. En esas películas los robots con inteligencia artificial dominaban el mundo. Sin embargo, en la práctica **las máquinas inteligentes no salían de los pocos laboratorios** que trabajaban en inteligencia artificial. Hubo una gran decepción y se difuminó el empuje inicial.

A pesar de esto y de otras muchas dificultades, primero como juguetes y más tarde como artefactos electrónicos o robots con algunas interesantes aplicaciones, **se fue avanzando**, aunque muy lentamente, en el campo de la inteligencia artificial.

Tanto es así que, allá por los **años 70**, la **RAND** Corporation encargó a **Hubert Dreyfus**, profesor de **filosofía** en la Universidad de Berkely, un estudio sobre las posibilidades de esa nueva ciencia. Lógicamente pretendía asesorarse sobre el tema antes de invertir en él. Aquel estudio terminó con la publicación de un libro, de nada menos que 250 páginas, que se hizo famoso. Se titulaba **“Lo que los ordenadores no pueden hacer. Una crítica de la razón artificial”** (Harper & Row, New York, **1972**).

En esencia ese libro, muy bien escrito y no tan bien razonado, constituía un claro **alegato** en contra de la inteligencia artificial. El punto más importante del libro era una afirmación que decía: **“La inteligencia humana y la artificial son esencialmente distintas, lo cual califica a la inteligencia humana como única e irrepetible”**.

Este libro levantó una **gran polvareda** entre los partidarios de la inteligencia artificial, que lo criticaron duramente diciendo que estaba escrito por un filósofo y que, por lo general, los filósofos siempre

estaban en contra de los avances de la ciencia porque, casi nunca los entendían.

En ese libro **Dreyfus afirmaba**, entre otras muchas ataques, que los programas que existen para jugar al ajedrez no demuestran que una **máquina electrónica pueda jugar al ajedrez**, porque lo que hace no es jugar sino sólo poner en marcha unos cuantos automatismos, pero que jugar es algo muy distinto, que consiste en pensar (analizar) que movimiento es el más conveniente entre los muchos posibles.

Unos años después, cuando ya los programas electrónicos de jugar al ajedrez eran mucho más potentes, **le prepararon una encerrona** al propio Dreyfus y le invitaron a jugar contra uno de esos programas. Dreyfus no era un jugador de ajedrez profesional pero si lo hacia al nivel que se conoce como "amateur". Pues bien, la partida, cuenta Herb Simon que fue su organizador, fue emocionante. Había mucho público y aunque no se cruzaron apuestas, según iba progresaba el combate, unos creían que ganaría Dreyfus y otros que lo haría la máquina. Finalmente, ante el sonrojo de Dreyfus, la máquina le dio mate. Algunos comentaron que Dreyfus muy despechado dijo: **"No creo que nunca un programa informático de ajedrez sea capaz de derrotar al campeón del mundo"**.

5. EL TEOREMA DE GÖDEL.

En medio de esta gran polémica entre partidarios y detractores de la inteligencia artificial, apareció un grupo a los que se les ocurrió **desempolvar unas ideas** desarrolladas por un famoso matemático, Kurt Gödel, posiblemente el mejor especialista en **lógica matemática** que haya existido, que había nacido, en los **primeros años del siglo** pasado, en la antigua **Checoslovaquia**.

En 1931 Gödel había publicado un trabajo en donde demostraba que: **"Por muy potente que sea un sistema lógico, siempre existirá la posibilidad de plantear proposiciones o enunciados que ni puedan ser probados ni puedan ser refutados dentro del propio sistema"**. A esta tesis se acostumbra a llamar el **Teorema de Gödel**.

Los científicos Ernest **Nagel** y James **Newman**, escribieron un libro (no muy extenso, 120 páginas) titulado **"La prueba de Gödel"** (New York University Press, 1958) en el que afirmaban apoyándose en este teorema, que **"es imposible construir una máquina que**

equipare en inteligencia el cerebro humano". Y en efecto esto es cierto, si por máquina se entiende lo que se entendía en aquellos tiempos. El quid de la cuestión está en que el citado teorema sólo es aplicable a los **sistemas cerrados**.

El teorema de Gödel no es sólo **una limitación para las máquinas sino que lo es también para los seres humanos**. Si a uno de estos se le impone la condición de ser un sistema cerrado, es decir si se le hace una cierta proposición y se le obliga a resolverla sin ayuda exterior, existirán muchos casos en los que no podrá hacerlo. Pero los seres humanos no son un sistema cerrado, sino que tienen la capacidad de recurrir a ayudas externas, tales como libros, profesores o consejeros, lo que les convierten en **sistemas abiertos** y por tanto **no les aplica el Teorema** de Gödel.

Todo lo anterior nos indica que si queremos construir máquinas (ordenadores digitales) que puedan simular la inteligencia humana, tendremos **que hacer que esas máquinas sean abiertas**, lo que equivale a decir que esas máquinas tendrán que **ser capaces de aprender**.

Esta es una conclusión muy importante, pero hay que constatar que con anterioridad muchos especialistas en inteligencia artificial ya habían llegado a ese mismo resultado, sin recurrir al Teorema de Gödel.

6. LA OPINIÓN DE UN CIBERNÉTICO.

Joseph **Weizenbaum**, profesor de **cibernética** en el Instituto Tecnológico de **Massachusetts**, se propuso intentar poner un poco de orden dentro de la confusión existente entorno la inteligencia artificial. El resultado de su esfuerzo fue **un magnífico libro** titulado "**Computer Power and Human Reason, from judgement to calculation**" (Freeman and Company, San Francisco, 1976). En mi modesta opinión, este es **el mejor libro** que se ha escrito, hasta la fecha, sobre las posibilidades de la inteligencia artificial.

Para comenzar Weizenbaum declara que su postura está del lado de los que creen **que debe haber límites** bien definidos respecto a lo que se **intente simular con los computadores**, frente a aquellos que opinan que los computadores **pueden, deben y harán cualquier cosa**.

Weizenbaum intenta aclarar lo que representan estas dos posturas, planteando dos preguntas muy importantes. La primera: **“¿Son todos los aspectos del pensamiento humano reducibles a formalismos lógicos?”**. Si lo fueran, no habría duda que los computadores podrían simular perfectamente ese pensamiento. Sin contestar, por el momento, a esa pregunta, enuncia la segunda, aún de mayor calado que la primera: **¿A qué categoría tecnológica pertenece la especie humana, en el supuesto de que pertenezca a alguna?**

En 1958, Simon y Newell, de la Universidad Carnegie-Mellon, hicieron la siguiente reflexión: **“Ahora existen en el mundo máquinas que piensan, aprenden y crean. Más aún, su habilidad para hacer estas cosas está creciendo rápidamente hasta que – en un futuro previsible – la variedad de los problemas que podrán manejar será equiparable a la variedad de los que maneja la mente humana”**. La opinión de Weizenbaum en relación con esta afirmación es **bastante escéptica** y, como ejemplos casi indiscutibles, afirma que nadie cree que en el futuro pueda haber computadores inteligentes que sean capaces de **escribir una partitura** de buena música o **pintar un cuadro** como los que están en los museos. Y para reforzar su opinión, explica que es indiscutible que **existen ideas que ninguna máquina podrá llegar a comprender, porque están relacionadas con objetivos que no son del mundo de las máquinas**. Resulta así que a una máquina le sería muy difícil comprender **la motivación para escribir una partitura** de las consideradas como geniales, que podrían ser intentar hacerse famoso, entusiasmar a los potenciales oyentes, ganar mucho dinero o cualquier otra aspiración personal.

Los seres humanos – de nuevo siglo con los argumentos de Weizenbaum - no son máquinas y aunque admito que esos seres **ciertamente procesan información, sin embargo no lo hacen de la manera que lo hacen los computadores**. En este caso, yo personalmente no estoy muy de acuerdo con el cibernético americano. No creo que eso sea un argumento en contra de la inteligencia artificial. Pienso que tampoco los aviones vuelan como lo hacen los pájaros sin embargo es **evidente que vuelan**.

Por otro lado, es cierto que los computadores electrónicos se enfrentan a un grave problema muy difícil de resolver como consecuencia de que su mundo es tan distinto al nuestro. Ellos no tienen **sentido común**. Weizenbaum cita un ejemplo, que ilustra muy bien esta enorme dificultad a la que tendrán que enfrentarse los

especialistas en inteligencia artificial: **una tradicional madre de familia**, dedicada por completo a las labores de su casa, aunque no haya ido a la universidad ni a la escuela superior, gracias principalmente a su enorme sentido común, resuelve a diario numerosos de problemas, que resultarían muy difíciles para una máquina.

En otra parte del libro, Wiezenbaum entra de pasada en esa afirmación tan simple, que argumentan tantas veces los opositores a la inteligencia artificial: **“Una máquina nunca podrá ser más inteligente que su programador”**. El cibernético explica que en la actualidad, tratándose de programas muy potentes, ya no existe la figura del programador tradicional que siempre aparece en el caso de los programas sencillos. Tratándose estos programas muy elaborados, ocurre que inevitablemente son modificados en infinidad de ocasiones, que se les añaden rutinas y subrutinas, así como toda clase de elementos auxiliares. El resultado final es que el tamaño del programa crece enormemente, gracias a lo cual es mucho más eficaz, sin embargo, la figura del programador ha desaparecido por completo, porque ya no habrá ninguno que pueda saber cómo se comportará el programa.

Aunque Wiezenbaum es muy escéptico en cuanto a la capacidad ilimitada de los ordenadores digitales, **su objeción más importante**, como ya se dijo anteriormente, se refiere al tema de **los objetivos** de las máquinas inteligentes. **“¿Qué objetivos y qué propósitos humanos no deberíamos delegar en las máquinas?”** se pregunta el cibernético americano. Para contestar a esta pregunta es necesario entrar en una larga serie de **consideraciones éticas**, que demuestran lo difícil que será para una máquina comprender esos conceptos. Pero por no alargarme demasiado, les voy a poner un **ejemplo, de mi propia cosecha**, pero muy de actualidad. Como se está alargando considerablemente la esperanza de vida de los españoles, resulta que cada vez hay más jubilados que cobramos una pensión de la seguridad social y menos trabajadores que cotizan a ella, hecho que está poniendo en serio peligro el aspecto económico del sistema actualmente establecido. Si planteáramos este problema a una máquina inteligente, ésta, después de analizar cuidadosamente toda la información que le suministráramos, sentenciaría: **“Eliminen, por el método más expeditivo, a todos los jubilados de más de ochenta años”**.

7. PRIMEROS ÉXITOS IMPORTANTES.

En **1997** se produjo un hecho inesperado y de una gran trascendencia en el ámbito de la inteligencia artificial. Una máquina digital, a la que habían puesto el nombre de **Deep Blue** y había sido fabricada por **IBM**, derrotó al campeón del mundo de ajedrez, **Gary Kaspárov**. La importancia del hecho radicaba en que, con anterioridad a esta fecha, todos los escépticos en inteligencia artificial habían pronosticado que **“eso nunca podría ocurrir”**.

Aunque fue una fecha memorable en el mundo de la inteligencia artificial, hay que dejar bien claro, que aquella máquina tenía aún **menos inteligencia que la de un mosquito**. Pero lo importante es que a partir de entonces se comprobó que acumular conocimiento en una máquina, formando un amplio conjunto de reglas sencillas, era el modo más práctico para que pudiera resolver problemas bastante complicados. Así empezaron a proliferar los llamados **sistemas expertos**.

Estos sistemas son ordenadores digitales bastante potentes, que **tienen codificada en su interior “la sabiduría y la experiencia del ser humano”** en un cierto campo, no muy amplio, de la vida cotidiana. Por ejemplo, el juego de ajedrez, el diagnóstico de algunas enfermedades, la resolución de problemas jurídicos, el análisis de espectrogramas, etc.

Posiblemente este campo de los sistemas expertos sea el que tiene **más futuro**, a corto plazo, dentro de la inteligencia artificial.

Aunque pueda parecer mentira, en **varios campos de la medicina**, estos sistemas expertos **superan a los mejores especialistas** en la materia que existen en el mundo. Esto, como ustedes pueden imaginarse, está teniendo enormes repercusiones en nuestra sociedad.

Los sistemas expertos, indudablemente **exhiben una cierta inteligencia**, aunque sea en un campo muy concreto y generalmente muy reducido. Sin embargo, yo creo, que conviene reflexionar sobre si son o no son inteligentes, porque si se piensa con detenimiento ellos **(en el caso concreto de un sistema experto que diagnostica enfermedades) trabajan exactamente igual a como lo hace un buen especialista en medicina**.

A diferencia de Deep Blue, que simplemente servía para jugar al ajedrez, por aquella época también se desarrolló otro sistema experto de un enorme interés científico. Se trataba del construido por el JPL de Pasadena, Centro Asociado de la NASA, cuyo primer objetivo fue **pilotar una sonda espacial** hasta la superficie del planeta Marte. Este sistema fue bautizado con el nombre **DEEP SPACE** y se instaló por primera vez en la sonda Mars Pahtfinder, lanzada al espacio en diciembre de 1996. El sistema llevaba un **micro ordenador IBM** y su trabajo consistía en controlar el despegue, guiar el rumbo de la sonda en la atmósfera marciana, mantener las comunicaciones con un pequeño robot todoterreno que se envió desde él a la superficie del planeta, manejar su cámara fotográfica panorámica, operar su estación meteorológica y algunas cosas más. DEEP SPACE, convenientemente adaptado algunas distintas necesidades, se utiliza actualmente en casi todas las sondas enviadas al espacio profundo por la NASA.

Cuando se habla de los progresos de la inteligencia artificial resulta necesario mencionar, aunque sólo sea de pasada, la irrupción en este campo de la **"Lógica Borrosa"** ("fuzzy logic", en inglés). Sabido es que en la mayoría de los problemas de control, se trabaja con mecanismos, ya sean físicos o electrónicos, que operan de la siguiente forma: **"Si** ocurre esto, **entonces** hay que responder con esto otro". En muchos casos este procedimiento falla porque en la práctica las variables "esto" y "esto otro", no aparecen bien definidas. Un ejemplo muy simple sería: "Si llueve, entonces hay que abrir el paraguas". Pero lo que suele ocurrir es que puede llover poquísimo, un poco, lo normal, bastante más, hasta caer un chaparrón o caer un chaparrón muy fuerte. Estos sistemas se caracterizan porque sus variables son **imprecisas** y han dado origen a **la ciencia de lo borroso**: conjuntos borrosos, tecnología borrosa, ingeniería borrosa, etc.. Todas estas disciplinas han dado origen a logros técnicos muy importantes en campos tales como los sistemas de control que, por ejemplo, funcionan con gran éxito en el metro de Tokio y de otras ciudades de Japón, en diversos ferrocarriles, en muchos equipos para el control del suministro de aguas, etc..

8. EL FUTURO A MEDIO PLAZO (del 2012 al 2050).

La asignatura pendiente más importante, en el campo de la inteligencia artificial, es conseguir que las máquinas digitales sean **capaces de aprender**. Hoy en día ya existen diversas máquinas que aprenden, sin embargo lo hacen de una manera muy rudimentaria. Lo

que se necesita es que haya máquinas digitales que sean capaces de aprender con bastante soltura, como lo hace un joven de 12 ó 15 años. Es muy arriesgado afirmar algo en este terreno, pero en mi modesta opinión, creo que **dentro de 10 ó 20 años se logrará este importante objetivo.**

Por otro lado, en las **Universidades de Stanford y en el MIT** se están desarrollando una serie de **robots modulares**, contruidos a base de módulos intercambiables que se montan como un lego. Cada uno de estos módulos tiene unas capacidades distintas lo que les permite adaptarse a necesidades muy variadas, la mayoría de ellas relacionadas **con trabajos domésticos, funciones de rescate, exploración en zonas peligrosas, auxilios en casos de incendios, etc..** Se cree que en 20 ó 30 años estos robots modulares podrán hacer **cosas difíciles de imaginar en la actualidad.**

Otro campo en el que se está **avanzando** de una manera insospechada, es el de los **robots cirujanos**. Se trata de sistemas expertos que permiten practicar **operaciones quirúrgicas** cada vez más delicadas **sin tener que abrir** al paciente, sino sólo mediante **minúsculas incisiones**. Se piensa que en un futuro no muy lejano ya nunca se **"cortará la piel"** (como suele decirse) de un enfermo, por muy grave que sea la enfermedad que padezca. La **cirugía no invasiva** será la norma en todos los hospitales. Más aún, en un futuro no muy lejano es posible que veamos robots cirujanos que sean capaces de **operar vasos sanguíneos microscópicos, fibras nerviosas y otros tejidos manipulando escalpelos, pinzas y agujas de tamaño microscópicos, cosas que ahora son totalmente imposibles.**

Aunque quizá les parezca a ustedes poco serio, hoy ya existen los **robots cocineros** capaces de elaborar los platos más comunes de **comida rápida** y no sería de extrañar que en un futuro no muy lejano proliferen cocineros digitales capaces preparar **otros platos mucho más elaborados.**

9. EL FUTURO A LARGO PLAZO (del 2050 al 3000).

A largo plazo, un avance que aún se resiste a los investigadores en inteligencia artificial, es desarrollar máquinas digitales que **posean consciencia**, es decir que no sólo sean capaces de procesar y elaborar información, sino que además, simultáneamente al hecho de procesar y elaborar, también sean capaces de **comprender el significado** de esa información que está procesando.

Esto de la consciencia, dicho así parece sencillo, sin embargo es **extremadamente complejo** para un ordenador digital. Contemplar la fotografía de un paisaje que contenga, por ejemplo, las murallas de Ávila con su entorno natural y poder comprender con la necesaria "coherencia" todo lo que aparece en ella, como lo hace en unos segundos un ser humano de 10 ó 12 años, es extremadamente complejo para una máquina. Si, por ejemplo, en lo alto de la muralla aparecen dos **vacas paciendo**, el humano inmediatamente se daría cuenta de que eso es imposible y por tanto que la foto está trucada. Sin embargo, una reacción como esa parece casi imposible para una máquina digital.

Anteriormente se ha dicho que hablar de algo imposible en ciencia es la mejor forma de equivocarse. A pesar de esto los investigadores en inteligencia artificial admiten que **aún estamos muy lejos de lograr la consciencia artificial.**

Puestos a imaginar, conviene reflexionar sobre lo que sucederá cuando los **robots llegan a ser más inteligentes que nosotros** mismos. Los científicos **no dudan de que esto sucederá**, sin embargo en lo que no se ponen de acuerdo es en el "**cuándo**". Ni tampoco en las **consecuencias que este acontecimiento**, tan temido por muchos y tan deseado por unos pocos, podrá originar.

Como estamos hablando de finales del siglo XXI, valdría la pena pensar: **¿Aprenderán los robots a ser humanos algún día?** Si se pregunta a los expertos en inteligencia artificial, la mayoría de ellos opinarían que eso **nunca sucederá**. No tiene mucho sentido pensar, por ejemplo, **en robots emotivos capaces de enamorarse**, o cosas por el estilo. Sin embargo, repetimos, **en ciencia decir que algo es imposible, es la mejor forma de equivocarse**, la historia así lo ha demostrado en muchas ocasiones.

Pensando en estas cosas, que quizá ocurran en el futuro, convendrá tener en cuenta que hay **una ley de la evolución** que establece que: "**las especies más aptas siempre desplacen a las menos aptas**". Esto viene a cuento porque podríamos crear una especie digital que pudiera llegar a desplazarnos. Ya lo decía **Wiezenbaum: "¡Cuidado con lo que delegamos en las máquinas!"**.

Para evitar que algún artefacto de hoja de lata nos desplace se está pensando que lo más conveniente sería crear eso que algunos

expertos han denominado: **“La Inteligencia Artificial Amistosa”**. Este término, acuñado por **Eliezer Yubkowsky del Instituto Singularidad para la Inteligencia Artificial**, se refiere a robots **diseñados para que deseen ayudar a los seres humanos**, en lugar de que deseen destruirlos. Es algo parecido a las **leyes de Asimov** de la robótica, pero **en lugar de imponer** la ley desde fuera hacia dentro del robot, se hace **desde dentro hacia fuera**. Esto ha creado un nuevo campo de la robótica y la inteligencia artificial, que llaman: **“La Robótica Social”**, en donde, aunque parezca mentira, ya hay **varios científicos trabajando**.

Sigamos viajando por los últimos años de nuestro siglo. **Rodney Brooks, antiguo director del famoso Laboratorio de Inteligencia Artificial del MIT**, afirma que entonces o aún bastante antes, se producirá un **cambio cultural** que nos llevará a **aceptar la introducción en nuestros cuerpos de tecnología robótica**, de silicio y acero, con el fin de **mejorar nuestras capacidades**. Este científico afirma que: **“Es posible que lleguemos a tener una conexión inalámbrica que nos conecte con internet instalada en nuestro cerebro”**. Otros expertos dudan que los beneficios de esta posibilidad superen algún día a los **riesgos de la cirugía cerebral**.

Podríamos seguir hablando de mundos fantásticos, por ejemplo, el de la **fusión de los humanos con los robots** y otras cosas por el estilo, pero **se nos está haciendo tarde**, así es que termino ahora, aunque antes me gustaría contarles una anécdota personal.

10. ANÉCDOTA PERSONAL.

Yo publiqué un libro titulado **“Cibernética de lo Humano”** en **1983**, aunque los escribí una par de años antes, porque como era mi primer libro, más allá de los publicados en el CSIC, me costó mucho que me lo publicaran.

En el año **1988** cayó en mis manos un libro escrito por **Juan Luis Ruiz de la Peña**, un catedrático de Teología de la Universidad Pontificia de Salamanca, titulado: **“Antropología Teológica Fundamental”**, en el que comentaba muchas partes de mi libro y, diciéndolo en plata, **“me ponía a parir”**, eso sí, con la elegancia propia de un teólogo y con la parafernalia de cientos de notas bibliográficas a pie de página propias de un ratón de biblioteca. Yo **no quise entrar en discusiones** con este buen señor, porque no quería perder el poco tiempo libre del que disponía, debido a mí trabajo con la NASA. Un par de años después este

teólogo publicó un segundo libro y también me ponía a parir. **Y luego un tercero** que no llegué a adquirir por falta de tiempo, pero que se me antoja, similar a los anteriores.

Cuando me **jubilé, en 1995**, como entonces ya disponía de mucho tiempo libre, decidí hablar con el autor de estos tres libros. Después de muchos esfuerzos conseguí localizarle, porque a pesar de ser catedrático en Salamanca, resulta que vivía en Oviedo. Cuando contestó a mi llamada telefónica y le dije mi nombre, lo primero que dijo fue: **iAtiza!**. Posteriormente nos hicimos bastante amigos, a pesar de nuestras discrepancias cibernéticas.

Recuerdo perfectamente como al teólogo le parecía inaceptable una frase que aparecía en mi libro: **"En el futuro los hombres serán esclavos de las máquinas"**, aunque con un **"Quizá"** delante, que el omitía en sus comentarios. Sin entrar en discusiones, que no me gustan, **le conté** como el año pasado, un avión de que volaba desde Venezuela a Madrid, con 389 pasajeros y 14 miembros de la tripulación, cuando ya enfilaba la pista del aeropuerto de Barajas, a unos 10 kilómetros de distancia, en medio de la oscuridad de la noche y de una niebla no muy densa, el computador de abordo le indicó al piloto: **"iPull up!"**, que en el lenguaje aeronáutico significa: "iLevanta el vuelo, vas muy bajo!". El piloto se quedó impávido y siguió volando hacia la pista de Barajas sin modificar su altura. El ordenador insistió "airado": "iiPull up!!" y entonces el piloto le contestó todavía más airado: **"iiCállate gringo!!"**. Esta conversación quedó registrada en las cajas negras del avión, porque unos segundos después se estrelló contra unos montículos y murieron los 389 pasajeros y los 14 tripulantes.

30 de abril de 2012
Revisado: 14 de junio de 2012
Penúltima revisión: 18 de junio de 2012
Última revisión: 23 de julio de 2012
Segunda última revisión: 22 del 9 de 2012

Filename: INTELIGENCIA ARTIFICIAL