

¿Por qué la gente piensa que los ordenadores no pueden [pensar]?

Por Marvin Minsky

Traducido por Wintermute

Original en

<http://www.ai.mit.edu/people/minsky/papers/ComputersCantThink.txt>

Publicado en su origen en el AI Magazine, vol.3 nº4, 1982. Reimpreso en Technology Review, Nov/Dic 1983, y en The Computer Culture, (Donelly, Ed.) Associated Univ. Presses, Cranbury NJ, 1985

Mucha gente piensa que los ordenadores nunca serán capaces de pensar. Quiere decirse, pensar de verdad. Ni ahora, ni nunca. También hay mucha gente de acuerdo en que los ordenadores pueden realizar muchas tareas para cuya realización una persona debería de pensar. Entonces, ¿cómo puede ser que una máquina aparente pensar pero que no piense realmente? Bueno, dejando aparte la cuestión acerca de qué es realmente pensar, creo que la mayoría de nosotros respondería que en esos casos el ordenador no está haciendo más que una imitación superficial de la inteligencia humana. Ha sido diseñado para obedecer determinadas órdenes sencillas, y después se le ha provisto de programas compuestos por estas órdenes. Debido a esto el ordenador ha de obedecer dichas órdenes, pero carece de idea alguna sobre lo que está sucediendo.

De hecho, cuando los ordenadores aparecieron la mayoría de sus diseñadores los fabricaron de cara a la realización de enormes computaciones carentes de "inteligencia". Por eso es que fueron llamados "computadores". Incluso entonces, algunos pioneros —especialmente Alan Turing— tuvieron la visión de lo que ahora llamamos "Inteligencia Artificial" —o "IA". Vieron que los ordenadores posiblemente podrían llegar más allá de la aritmética, y quizá imitar los procesos presentes en el cerebro humano.

Hoy en día, con la invasión de robots en la industria y en las películas, la mayor parte de la gente piensa que la IA ha llegado mucho más lejos de lo que realmente lo ha hecho. Y aun así, los "expertos" de hoy en día dicen que las máquinas nunca pensarán realmente. Entonces, ¿cómo es posible que puedan llegar a ser tan inteligentes y al tiempo tan estúpidas?

¿Puede ser creativa una máquina?

Naturalmente, admiramos a nuestros einsteins y beethovens, y nos preguntamos si un ordenador sería capaz de producir teorías —o sinfonías— tan magníficas. La mayor parte de la gente opina que la creatividad requiere algún tipo de "regalo" mágico, un talento especial que no puede ser explicado. Si así fuera, ningún ordenador podría crear - puesto que todo lo que una máquina pueda hacer (piensa la mayoría) puede ser explicado.

Para ver qué hay de erróneo con esto, debemos evitar cierta trampa para ingenuos: no debemos tan sólo centrarnos en trabajos que nuestra cultura admita como grandes, sino que antes hemos de aprender acerca de cómo la gente ordinaria hace cosas normales. No podemos esperar adivinar cómo los grandes compositores escriben grandes sinfonías. No creo que haya una gran diferencia entre el pensamiento ordinario y el altamente creativo. No culpo a nadie por no poder hacer lo que hace la gente más creativa. No les culpo por no ser capaces de explicarlo, tampoco. A lo que me opongo es a la idea de que, simplemente por ser incapaces de explicarlo ahora, no vayamos a poder entender nunca como funciona la creatividad.

No debería de intimidarnos la admiración hacia nuestros beethovens y einsteins. En lugar de ello, deberíamos de sorprendernos sobre nuestra ignorancia acerca de la forma en que obtenemos las ideas –y no me refiero tan sólo a las "creativas". Estamos tan acostumbrados a la adoración de lo inusual que olvidamos lo poco que conocemos sobre las maravillas del pensamiento ordinario. Quizá nuestras supersticiones sobre la creatividad podrían ser útiles de otro modo, como por ejemplo admirando héroes con cualidades tales que, de algún modo, nuestras deficiencias parezcan más perdonables.

¿Hay alguna diferencia especial entre aquellas mentes que destacan y el resto? No creo que exista nada básicamente diferente en un genio, excepto una combinación de habilidades de las cuales –curiosamente– ninguna en particular es muy especial. Ha de existir una vocación intensa respecto de algún tema, pero eso desde luego es muy común. También ha de existir una gran habilidad con éste tema, lo cual, tampoco es tan raro; solemos llamarlo "artesanía". Ha de haber también la suficiente confianza en uno mismo como para poder mantener el tipo; llamamos a esto tenacidad. Ciertamente, también debe haber sentido común. Como yo lo veo, cualquier persona normal que pueda entender una conversación normal tiene ya en su cabeza la mayor parte de lo que nuestros héroes tienen. Así pues, ¿cómo puede el sentido común ordinario –bien ordenado y ante todo muy motivado– hacer de alguien un genio?

Así que aún debemos preguntarnos, ¿por qué no todo el mundo adquiere una combinación así? En primer lugar, a veces es simplemente que se da el accidente de encontrar una nueva forma de ver las cosas. Pero aún entonces, deben haber algunas diferencias de grado respecto a otros. Una está en cómo este tipo de gente aprende a manejar aquello que aprende: más allá de su maestría aparente, las personas creativas han de poseer una serie de habilidades inconscientes de administración del conocimiento que atan juntas las cosas que saben. La otra diferencia reside en por qué algunas personas aprenden muchas y mejores habilidades. Un buen compositor domina gran cantidad de habilidades expresivas –pero también lo hace alguien que habla de forma coherente.

¿Por qué hay gente que aprende tanto tan rápido? ¡La hipótesis más simple es que han encontrado mejores caminos para aprender! Quizá tales "regalos" son poco más que pequeños trucos. Así como un niño aprende a reconstruir edificios con bloques mediante caminos inteligentes, ¡otro niño podría jugar dentro de su cabeza a reordenar la forma en que aprende!

Nuestra cultura no nos anima a pensar mucho acerca del aprendizaje. En lugar de ello, lo consideramos como algo que sucede, sin más. Pero el proceso del aprendizaje ha de consistir en sí mismo de toda una serie de habilidades que nosotros desarrollamos; comenzamos con tan sólo algunas de ellas, y poco a poco el resto van haciendo su aparición. ¿Por qué no se hace hincapié en esta adquisición de nuevos y mejores métodos de aprendizaje? Esto se debe a que no hay una recompensa directa, a que los resultados no son inmediatos. Cuando los niños juegan con cubos y arena, sus objetivos son cosas como llenar los cubos con la arena. Pero una vez un niño se dedica al objetivo de cómo aprender mejor, ¡podría llegar a una curva de aprendizaje exponencial! Cada nuevo camino para aprender a aprender llevaría a mejores formas de hacerlo –y esto podría ampliarse hasta el punto de ser un gran cambio cualitativo. Así pues, la gran creatividad podría ser la consecuencia de pequeñas coincidencias infantiles.

Así pues, ¿por qué es tan raro el genio, si todos tenemos lo que se necesita? Quizá porque nuestra evolución se basa en un desprecio sin sentido del individuo. Estoy seguro de que ninguna cultura podría sobrevivir si cada uno encontrara formas diferentes de pensar. Qué triste si así fuera, puesto que los genes del genio habrían de ser reprimidos, en lugar de desarrollados.

Resolución de problemas

Difícilmente vamos a encontrar la forma de hacer que una máquina haga maravillas si no averiguamos el modo de hacer las cosas del día a día. Los primeros programas de ordenador eran poco más que listas y bucles de comandos del tipo "Haz esto. Haz lo otro. Haz esto y aquello hasta que tal cosa suceda". La mayor parte de la gente aún escribe programas en este tipo de lenguaje (como BASIC o FORTRAN), que te fuerzan a imaginarte lo que el programa hará en cada momento. Llamemos a esto, la programación tipo "Haz Ahora".

No pasó mucho tiempo hasta que los investigadores en IA encontraron nuevos caminos para hacer programas. En el sistema "General Problem Solver" (resolutor general de problemas), escrito a finales de los cincuenta, Allen Newell, J. C. Shaw y Herbert A. Simon mostraron formas en las que describir procesos en forma de afirmaciones como "si la diferencia entre lo que tienes y lo que quieres es del tipo D, intenta encontrar esa diferencia utilizando el método M". Esto es lo que llamamos métodos de programación del tipo "Haz si es Necesario". Este tipo de programas aplican reglas de forma automática cada vez que se las necesita, por lo que los programadores no han de anticipar lo que va a suceder. Esto, inició una nueva ola de programas que podían resolver problemas de formas en las que el programador no había pensado, puesto que se podía decir qué tipo de cosas habían de probar sin saber de antemano qué método funcionaría. Todo el mundo sabe que si probamos suficientes veces al azar, al final seremos capaces de hacer algo. Pero cuando esto sucede en trillones de años, como nosotros, monos aporreando teclas al azar en un teclado mecánico, no se llama inteligencia –se llama evolución. Los nuevos sistemas no hacían las cosas al azar, pero utilizaban "consejos" sobre qué es probable que funcionase en un tipo determinado de problema. Así pues, en lugar de intentar adivinar al azar, estos programas podían tantear alrededor, del mismo modo en que escalarías una montaña en la oscuridad, tanteando. El único problema fue la tendencia a quedarse atascados en picos más pequeños, en lugar de llegar a la auténtica cumbre de la montaña.

Desde entonces, mucha de la investigación en IA ha sido orientada a la búsqueda de métodos más "globales" para evitar estos atascos, mediante la fabricación de programas que planifiquen de antemano y tengan una visión más amplia del problema. Aún, nadie ha descubierto un camino "totalmente general" para encontrar siempre el mejor método –y tampoco nadie lo espera.

En su lugar, hoy en día muchos investigadores en IA dirigen sus esfuerzos a programas que busquen patrones en memoria para decidir qué hacer después. Me gusta llamar a este tipo de programación el tipo "Haz Algo Sensato". Unos pocos investigadores –demasiado pocos, en mi opinión– experimentan con programas que pueden aprender y razonar mediante analogías. Estos programas algún día recordarán cuáles de sus viejas experiencias están más cercanas a una nueva situación, pudiendo recordar el método que mejor funcionó en problemas similares en el pasado.

¿Pueden un ordenador entender?

¿Podemos hacer que un ordenador entienda lo que le decimos? En 1965, Daniel Bobrow escribió uno de los primeros Sistemas Expertos basados en reglas. Se llamaba "STUDENT", y era capaz de resolver una amplia variedad de problemas de álgebra al estilo de los siguientes:

La distancia entre Nueva York y Los Ángeles es de 3000 millas. Si la velocidad media de un avión es de 600 millas por hora, ¿cuánto tiempo tardará en llegar desde Nueva York a Los Ángeles?

El tío del padre de Bill es el doble de viejo que el padre de Bill. En dos años, el padre de Bill será tres veces más mayor que Bill. La suma de sus edades es de 92. ¿Cuál es la edad de Bill?

La mayor parte de los estudiantes encuentran estos problemas mucho más difíciles de resolver que las ecuaciones formales algebraicas que les enseñan. Resolver estas ecuaciones parte de una receta –pero para resolver problemas informales basados en palabras, tienes que plantear las ecuaciones a resolver, y para ello has de entender lo que las palabras y frases quieren decir. ¿Entendía STUDENT? No. En su lugar, utilizaba unos cuantos trucos. Fue programado para adivinar que la palabra "es" normalmente significa "es igual a". Tampoco intentaba averiguar qué significaría "El tío del padre de Bill" –tan sólo se daba cuenta de que esta frase se refería al "padre de Bill". No sabía que palabras como "edad" y "mayor" hacen referencia al tiempo, pero los utilizaba para representar números de cara a las ecuaciones. Con unos cuantos trucos de este estilo, STUDENT a veces conseguía obtener la respuesta correcta.

¿Diríamos entonces que STUDENT "entendía" estas palabras? Para qué preocuparse. ¿Por qué caer en la trampa de sentir que hemos de definir viejas palabras como "significado" y "entender"? Resulta magnífico cuando las palabras nos ayudan a tener buenas ideas, pero no cuando nos confunden. La pregunta debería de ser: ¿evita STUDENT los "verdaderos significados" utilizando trucos?

O... ¿es quizá que lo que llamamos significados son en realidad grandes colecciones de trucos? Tomemos un ejemplo clásico, como el del significado de los números. Obviamente, STUDENT sabe algo de aritmética, en el sentido de que puede encontrar respuesta a cosas como "5 mas 7 es 12". Pero, ¿entiende los números en algún otro sentido –por ejemplo, ¿qué es "5"?– o qué significan "más" y "es"? ¿Que dirías tú si te preguntase "qué es cinco"? A principios de este siglo, los filósofos Bertrand Russell y Alfred North Whitehead propusieron una nueva forma de definir los números. "Cinco", decían, es "el conjunto de todos los posibles conjuntos de cinco miembros". Este conjunto incluye cada conjunto de cinco bolígrafos y cada camada de cinco gatitos. Desafortunadamente, también incluye conjuntos como "las cinco cosas que menos esperarías" o "los cinco números más pequeños no incluidos en este conjunto"– y esto llevó inevitablemente a inconsistencias y paradojas. El objetivo básico era encontrar definiciones perfectas para las ideas y palabras de uso ordinario. Pero si incluso al intentar hacer funcionar esta idea en las Matemáticas se dan estas inconsistencias, la teoría de Russell-Whitehead se hace demasiado compleja como para poder utilizarla de un modo práctico. Los educadores pertenecientes al movimiento de las "Nuevas Matemáticas" en los 60 intentaron hacer que los niños usaran esta teoría de conjuntos; tan sólo se consiguió separar aún más a aquellos a quienes les gustaban las matemáticas de aquellos que las despreciaban. Creo, que el problema estaba en que se ignoró un hecho esencial de la mente: lo que algo significa para mí depende hasta cierto grado de las otras cosas que sé.

¿Qué sucedería si construyéramos máquinas no basadas en definiciones rígidas? ¿No acabarían hundiéndose en la paradoja, en la inconsistencia? ¡Relax! La mayoría de lo que la gente "sabemos" está repleto de contradicciones, y aun así seguimos vivos. Lo mejor que podemos hacer es ser razonablemente cuidadosos; dejemos que nuestras máquinas también lo sean. Si aun así es posible que se produzca un error, pues qué le vamos a hacer, la vida es así.

Redes de conocimiento

Si cada significado en la mente depende de otros significados en la misma mente, ¿lo convierte eso en un indefinible, al menos de cara al trabajo científico? No, ¡incluso cuando el concepto se mueve en círculos hay ciencia esperándonos! Sencillamente, hacer nuevos tipos de teorías, ¡sobre los propios círculos! Las viejas teorías sólo evitaban los círculos. Pero eso hizo que perdieran toda la riqueza de las redes de significados humanas; las redes presentes en nuestro cerebro son probablemente la estructura más compleja que la ciencia jamás haya observado. Del mismo modo, las teorías en IA desarrolladas al detalle es probable que necesiten a su vez teoría muy compleja. Pero así es la vida.

Volvamos al significado de los números. Esta vez, para hacer las cosas más fáciles, pensaremos sobre el Tres. Estoy defendiendo que el Tres, para nosotros, no tiene una definición sencilla; es, más bien, una red de diferentes procesos los cuales obtienen el significado de los demás. Considérese todos los papeles que "Tres" juega. Una forma de definir un Tres es recitar "Uno, Dos, Tres", mientras señalamos a cosas diferentes. Para hacerlo correctamente se debe de (i) tocar cada cosa una sola vez y (ii) no tocar ninguna dos veces. Otra forma es contar en voz alta mientras cogemos cada objeto y lo volvemos a dejar. Los niños aprenden a hacer este tipo de cosas mentalmente o, cuando es muy complejo, mediante pequeños trucos como aquel de señalar las cosas. Otra forma de mostrar lo que es Tres es utilizar algún conjunto estándar de Tres cosas. Entonces, llevar ese conjunto a otro conjunto, y relacionarlos uno a uno: si todos están emparejados y no queda ninguno, entonces seguro que había Tres. Ese conjunto de Tres no necesita ser concreto, no se necesita que sean "cosas"; las palabras "uno, dos, tres", funcionan igual de bien. Para Cinco, tenemos un rango de elección más amplio. Uno puede pensar en ello como grupos de Dos y Tres, o de Uno y Cuatro. O incluso, uno puede recordar figuras familiares como la de un pentágono, una X, una V, una cruz o un aeroplano; todas ellas forman un Cinco:

o o	o o	o o	o o
o o	o	o o	o o o o o o
o	o o	o	o o

Dado que cada truco trabaja sobre una situación diferente, nuestro poder reside en la capacidad de cambiar de un truco a otro. Preguntar cuál es el significado correcto –contar, enlazar, o hacer un grupo– es absurdo. Cada uno de ellos tiene su utilidad y relación con los demás. Ninguno de ellos tiene un gran poder por sí mismo, pero juntos conforman un sistema de habilidades bastante versátil. En lugar de formar una endeble cadena enlazada de definiciones en nuestra mente, cada palabra que utilizamos puede activar grandes redes que contengan caminos diferentes para referirse a las cosas, para utilizarlas, para recordarlas, para compararlas, etc. No podemos quedarnos atascados si poseemos redes de conocimiento con conexiones múltiples. Cuando un sentido del significado falla, podemos cambiar a otro. Por el camino del matemático, en cuanto te encuentres con el más mínimo problema, estarás atascado sin remedio.

¿Por qué entonces los matemáticos se ajustan a cadenas tan limitadas, haciéndolas depender del menor número posible de cosas? La respuesta tiene su ironía: ¡los matemáticos quieren atascarse! Para ellos la fragilidad no es mala, pues les ayuda a encontrar la prueba perfecta, no sea que vayan a encontrar cosas que sean inconsistentes entre sí. Es una buena política para los Matemáticos; de hecho, en esto consiste gran parte de las matemáticas. Sin embargo, no es una buena Psicología. Afrontémoslo, nuestras mentes siempre tendrán creencias que van a resultar ser falsas.

Creo que es una mala psicología, cuando los profesores ajustan las matemáticas de nuestros niños a una frágil torre encadenada de largas definiciones, en lugar de a robustas redes interconectadas. Estas cadenas se rompen en sus eslabones más débiles, las torres se caen ante el más ligero empujón. Y eso es lo que sucede en la mente de un niño que atiende una clase de matemáticas, el cual gasta tan sólo un breve momento en ver una bonita nube pasar. Los propósitos de la gente normal no son los mismos que los de los matemáticos y los filósofos, que pretenden simplificar sus teorías reduciendo las conexiones a las mínimas que necesiten. En la vida real, las mejores ideas son aquellas cuya densidad de conexiones es mayor. Quizá es por ello que en nuestra cultura los niños tienen tanto terror a las matemáticas. Pensamos que les estamos ayudando a hacer bien las cosas, ¡haciendo que la mayor parte del tiempo vayan mal! Es posible que en lugar de enseñarles así, debiéramos ayudarles a construir redes más robustas dentro de sus cabezas.

Castillos en el aire

El secreto del significado de los conceptos reside en los caminos a través de los cuales están conectados al resto de las cosas que sabemos. A más enlaces, más significado va a tener una cosa para nosotros. Lo divertido viene cuando alguien pretende buscar el significado "real" de algo, dado que, en caso de que sólo tuviera un significado, esto es, si solo estuviera conectado a otra cosa –y sólo una–, ¡entonces no significaría nada en absoluto!

Por eso no creo que debamos programar nuestras máquinas de esa forma, utilizando definiciones lógicas y claras. Una máquina programada de esa manera nunca legará a entender nada –no como una persona podría. Una red provista de múltiples conexiones proporciona los suficientes caminos para utilizar el conocimiento como para que si uno de los caminos no funciona, podamos preguntarnos por qué. Cuando poseemos suficientes significados en nuestra red, podemos dar vueltas a las cosas en la cabeza y observarlas desde una perspectiva diferente cuando nos hayamos quedado atascados, pudiendo seguir entonces otro camino. ¡A eso nos referimos cuando decimos "pensar"!

Por eso es por lo que no me gusta la lógica y prefiero trabajar con redes de definiciones circulares. Cada individuo da significado al resto. No hay ningún problema porque nos gusten canciones diferentes y cada una de ellas sea a la vez nuestra favorita por la forma en que contrasta con las demás. No hay nada de malo en las cuerdas –o en un nudo, o en la ropa tejida a mano - donde cada hebra ayuda a sostener las demás– ¡o a mantenerlas separadas! ¡No hay nada erróneo en que, recorriendo este curioso camino a la hora de ver las cosas, la totalidad de nuestra mente sea un castillo en el aire!

Resumiendo: por supuesto que un ordenador no puede entender nada que sea real –ni tampoco qué es un número– si se le fuerza a manejarlo por un único camino. Pero tampoco podría un niño, ni un filósofo. Tales disquisiciones no tienen nada que ver con los ordenadores, sino que vienen dadas por nuestra absurda búsqueda de significados que "sean por sí mismos", fuera de todo contexto. Nuestras preguntas acerca de las máquinas inteligentes deberían de plantearse como preguntas sobre nuestras propias mentes.

¿Son los humanos conscientes de sí mismos?

La mayor parte de la gente asume que un ordenador no puede ser consciente de sí mismo; que como mucho, pueden simular que lo son. Por supuesto, esto está basado en la suposición de que los propios humanos somos conscientes de nosotros mismos. Pero, ¿lo somos? No lo creo. Sé que suena ridículo, así que me explicaré.

Si por consciencia entendemos "saber que es lo que contiene nuestra mente", entonces, como todo psicólogo clínico sabe, la gente es más bien poco consciente de sí misma, y la mayor parte de lo que saben de sí mismos es pura adivinación. Construimos auténticas redes de teorías sobre lo que tenemos en la cabeza, y confundimos esa visión aparente con lo que realmente está sucediendo. A las claras, la mayor parte de lo que nos revela nuestra "consciencia" es algo que nosotros mismos hemos preparado. No quiero decir que no seamos conscientes de los sonidos o de lo que nos transmite el sentido de la vista, o incluso de parte de nuestros pensamientos. Simplemente, digo que no somos conscientes de una buena parte de lo que sucede en nuestras mentes.

Cuando alguien habla, la explicación física está bastante clara: la voz hace temblar el aire, esto hace que nuestros receptores auditivos se muevan, y después los computadores de nuestra cabeza transforman estas ondas en los constituyentes de las palabras. Estos se transforman de algún modo en cadenas de símbolos que representan palabras, con lo que hay entonces algún lugar en nuestra cabeza que "representa" una frase. ¿Qué sucede después?

Cuando la luz excita nuestras retinas, esto activa toda una serie de eventos en el cerebro referentes a texturas, bordes en los objetos, colores y demás. Entonces, todo esto se fusiona para "representar" una figura, la forma de alguna cosa... de nuevo, ¿qué sucede después?

Todos entendemos estas ideas. Pero aún nos queda un grave problema. ¿Qué entidad, qué mecanismo continúa el proceso en el punto en que lo dejamos? Estamos acostumbrados a decir, sencillamente, que esta entidad es el "yo". ¿Qué hay de equivocado en esta idea? Nuestro concepto estandarizado del "yo" es que dentro de la mente reside algún tipo de "yo" central que es el que realmente hace el trabajo para nosotros, una pequeña persona que está ahí escondida para escuchar y ver y entender lo que está pasando. Llamamos a esto la teoría del "Agente Único", y no es difícil entender por qué toda cultura se adscribe a esta teoría. No importa lo ridículo que parezca científicamente; al contrario, se trata de la base de todos nuestros principios sobre ley, trabajo y moralidad. Sin esta teoría, todas nuestras afirmaciones acerca de las responsabilidades se derrumbarían; la culpa y la virtud, lo correcto y lo equivocado... ¿De qué nos valdría resolver problemas sin ese mito, incluso, cómo podríamos tener tan siquiera sociedades sin él?

El problema es que no podemos construir ninguna teoría mental coherente sobre este supuestos. En todo campo científico, estamos forzados a reconocer que lo que vemos como "una sola cosa" —como pueda ser una roca o una nube... o una mente— ha de ser descrito en ocasiones como algo formado por otro tipo de cosas. Necesitamos entender de alguna forma que el "yo", no es una sola cosa.

Nuevas teorías sobre la mente y las máquinas

Es muy fácil decir cosas como que "Los ordenadores no pueden hacer (xxx) porque no tienen sentimientos/pensamientos". Pero veamos una forma de hacer que esas afirmaciones suenen estúpidas. Cambiémoslas para que digan algo como esto: "los ordenadores no pueden hacer (xxx), porque lo único que hacen es ejecutar procesos increíblemente intrincados, quizá millones al mismo tiempo". Tales objeciones suenan ya bastante menos convincentes —y lo que hicimos fue bastante simple, aunque desvela una realidad compleja: aún no sabemos cuáles son los límites de la informática. Enfrentémonos a otra realidad igual de simple: nuestras nociones sobre la mente humana son tan primitivas como las que tenemos sobre la informática.

¿Por qué nos da tanto miedo admitir lo poco que conocemos sobre cómo funciona la mente humana? Debe de venir en parte de nuestra habitual tendencia a reprimir aquellos problemas que parecen descorazonadores. Pero hay razones más profundas, también, para creer en la unicidad e inexplicabilidad del "yo". Quizá tenemos miedo a que demasiadas preguntas aparten los velos que encumbran nuestra vida mental.

Para mí resulta especialmente irónico cuando la gente dice que las máquinas no pueden tener una mente, dado que veo como sólo estamos empezando a ver cómo podría funcionar una mente —todo esto mediante introspecciones provenientes de nuestros intentos de averiguar lo que una máquina compleja puede hacer. Por supuesto, estamos aún alejados de una teoría clara y completa. Pero ahora resulta extraño pensar que alguien pudiera haber esperado entender este tipo de cosas antes de haber averiguado bastante más sobre las máquinas. Excepto, claro está, si creían que la mente no es en absoluto un hecho complejo.

Ahora bien, supongo que la pregunta es, ¿si el concepto de "yo" está tan equivocado, que recomendarías en su lugar? Para empezar, no recomiendo cambiar nada en el ámbito social —es demasiado arriesgado. Pero si queremos avanzar en nuestra empresa dedicada al desarrollo de máquinas inteligentes, necesitamos teorías mejores acerca de como "representar", dentro de un ordenador, las redes de conocimiento y demás mecanismos que figuran en nuestras estructuras de

conocimiento. Debemos desarrollar programas que sepan y nos digan lo que significa un número, en lugar de reducirse a sumarlos y restarlos. Debemos experimentar con todo tipo de conocimiento general, y a su vez con el conocimiento sobre éste.

Éste, es el enfoque de parte de la investigación hoy en Inteligencia Artificial. Ciertamente, gran parte del mundo de la informática se dedica a construir programas grandes y útiles al tiempo que superficiales. Pero algunos estudiantes atrevidos intentan hacer que los ordenadores desarrollen otras formas de conocimiento, representando diferentes tipos de conocimiento, con el objetivo de que utilizando diferentes caminos sus programas no se queden atascados ante sus propias ideas inmutables. Lo más importante, quizá, es hacer que tales máquinas aprendan de su propia experiencia. Una vez sepamos más sobre estas cosas, podemos explorar caminos para unir estos esfuerzos. Finalmente, tendremos máquinas que piensen sobre sí mismas y hagan teorías, buenas o malas, sobre cómo funcionan. Quizá, cuando nuestras máquinas lleguen hasta este punto, nos demos cuenta de que resulta muy fácil decir lo que ha sucedido. Probablemente, en ese punto ellas no quieran considerarse máquinas. Aceptar eso será difícil, pero este sacrificio será la única forma de que las máquinas nos liberen de nuestras falsas ideas.

Conocimiento y sentido común

Sí, todos nos hemos divertido con bromas acerca de lo estúpidas que son las máquinas y cómo entienden las cosas de forma literal. Nos envían cheques absurdos con facturas de cero dólares. No pueden distinguir entre un guión y el símbolo de la resta. No les importa entrar en bucles infinitos, haciendo lo mismo una y mil veces. Esta falta absoluta de sentido común es otra razón por la que se piensa que no pueden existir máquinas que posean una mente. No es ya que sólo hagan lo que se les dice, es que son tan estúpidas que es casi imposible indicarles cómo hacerlas bien.

¿No resulta extraño pensar que incluso los primeros programas en IA eran auténticos genios en campos "avanzados", aun careciendo de sentido común? En 1961, un programa escrito por James Slagle podía resolver problemas de cálculo al nivel de un estudiante de instituto; incluso obtuvo una A en un examen del MIT. Sin embargo, no fue hasta 1970 que pudimos construir un programa-robot que podía ver y moverse lo suficientemente bien como para utilizar los bloques de construcción de los niños pequeños y hacer cosas como apilarlos, derribarlos, reordenarlos y meterlos en cajas.

¿Por qué pudimos construir programas con un conocimiento tan "experto" antes de construir otros con habilidades tan infantiles? La respuesta es una paradoja inesperada: ¡gran parte del conocimiento "experto" de los adultos resulta ser mucho más simple de lo que sucede cuando un niño juega! ¡Puede ser más difícil ser un novato que un experto! Esto se debe a que en ocasiones lo que un experto necesita saber y hacer es bastante simple –tan sólo, que puede ser muy difícil de descubrir o aprender en primer lugar. Galileo debió de ser muy astuto para darse cuenta de la necesidad del cálculo. No pretendió inventarlo. Hoy, cualquier buen estudiante puede aprenderlo.

Lo sorprendente es que cuando fue terminado el programa de Slagle sólo necesitó cerca de cien "sentencias" para resolver toda esta serie de problemas de cálculo con un nivel de instituto. Muchos de ellos no eran más que sencillas reglas sobre álgebra. Pero otros trataban sobre cómo adivinar cuál dentro de dos problemas podía resultar más sencillo; este tipo de conocimiento es de especial importancia, puesto que ayuda al programa a hacer juicios correctos acerca de cuál es el siguiente paso a seguir. Sin este tipo de conocimiento, este tipo de programa pierde su valor; con él, sus objetivos se alcanzan con mucha más facilidad. ¿Por qué tardan tanto los estudiantes humanos en aprender todas estas reglas? No lo sabemos.

Hoy en día sabemos mucho más sobre cómo hacer este tipo de programas "expertos" –pero no mucho más sobre cómo hacer programas con más "sentido común". Consideremos la cantidad de cosas que hace un niño cuando juega con sus bloques. Para construir una casita, el niño tiene que mezclar y utilizar buena cantidad de tipos distintos de conocimiento: sobre formas y colores, espacio y tiempo, gravedad, balance, presión, velocidad, coste... En su lugar, un experto puede seguir adelante con un conocimiento mucho más reducido; técnicamente, el sentido común es mucho más complejo.

La mayor parte de los programas informáticos no hacen más que aquello para lo que han sido programados. Algunos programas en IA son más flexibles; cuando algo va mal, regresan al último punto en que tomaron una decisión y prueban algo distinto. Pero esto sigue siendo una base muy cruda para la inteligencia. Para hacerlos realmente listos, tenemos que hacerlos más reflexivos. Cuando algo va mal las personas intentamos entender qué es lo que falla en lugar de dedicarnos a probar otro camino. Buscamos las causas –o las excusas–, y cuando las encontramos, las añadimos a nuestras redes de creencias y entendimiento. Hacemos un aprendizaje inteligente. Algún día, los programas podrán hacer esto también –pero aún necesitamos investigar mucho para averiguar cómo.

Miedos inconscientes y fobias

Apuesto a que cuando consigamos hacer máquinas más sensibles, nos daremos cuenta de que aprender lo que está mal es tan importante como aprender lo que es correcto. Para tener éxito, ayuda conocer los caminos del fracaso. Freud habló sobre los "censores de la mente", que nos protegen de actos y pensamientos prohibidos. Y aunque estos censores tienen el propósito original de regular nuestra actividad social, creo que también los utilizamos habitualmente a la hora de resolver problemas –hemos de saber qué es lo que no debemos hacer. Quizá cada vez que metemos la pata aprendemos algo nuevo, construyendo en algún tipo de "memoria subconsciente" un proceso para reconocer circunstancias similares.

Esta no es una idea muy popular en la psicología contemporánea, puesto que los censores, al limitarse su labor a la supresión de una conducta, realizan una actividad que resulta invisible en la superficie. Cuando alguien toma una buena decisión, tendemos a preguntarnos por la "línea de pensamiento" en que se ha basado. Pero no solemos preguntar con tanta frecuencia sobre las mil prohibiciones que han evitado mil alternativas erróneas. Si los censores trabajan dentro de nuestras mentes para evitar que caigamos en el error, ¿por qué no nos damos cuenta de que está sucediendo? Supongo, que se debe a que hay tantos miles de ellos funcionando a la vez que si tuviéramos que pensar sobre ellos no haríamos muchas cosas. Deben de protegernos de las malas ideas antes de que pensemos en utilizarlas.

Quizá esta es la razón por la que gran parte del pensamiento humano es "inconsciente". Cada idea que tenemos la oportunidad de contemplar es el producto de una buena cantidad de eventos que tienen lugar anteriormente –y más profundamente– en nuestro cerebro. Cada pensamiento consciente ha de ser el final de un proceso en el que debe competir con otros proto-pensamientos, quizá mediante pequeños contenciosos en minúsculos juzgados; pero aquello de lo que acabamos siendo conscientes no son más que las sentencias definitivas.

¿Y cómo podría ser de otra manera? No hay forma en la que una parte de la mente pueda saber todo lo que está sucediendo en el resto. Nuestras mentes conscientes deben actuar como altos ejecutivos, que no pueden distraerse con los pequeños detalles. Sólo hay tiempo para los resúmenes escritos por otras partes más pequeñas de la mente, que saben mucho más sobre mucho menos; estas, realizan el auténtico trabajo.

Ordenadores conscientes de sí mismos

Bien, ¿es posible programar un ordenador de modo que sea consciente de sí mismo? Normalmente se espera que la respuesta sea "no". ¿Y si respondiésemos que las máquinas son capaces, en principio, de una consciencia bastante mejor de la que tiene el ser humano?

Creo que esto podría hacerse proveyendo a las máquinas de formas de examinar sus propios mecanismos mientras trabajan. En principio esto parece ser posible; ya tenemos algunos programas sencillos capaces de entender un poco acerca de cómo funcionan programas más sencillos que ellos. (Hay un problema técnico al hacer a un programa lo suficientemente rápido como para alcanzarse a sí mismo, pero se puede solucionar registrando y almacenando su actividad). El problema reside en que aún sabemos demasiado poco como para hacer programas con el sentido común necesario para entender el funcionamiento del más simple de los programas de IA que resuelven problemas. Pero una vez aprendamos a fabricar máquinas lo bastante inteligentes como para entender este tipo de cosas, no veo ningún gran problema para darles la "introspección" necesaria para entenderse, cambiar y mejorarse.

Puede que esta no sea una solución muy sabia. ¿Pero qué sucedería si resulta que el único camino para hacer más inteligente a un ordenador es hacerlo más consciente de sí mismo? Por ejemplo, puede que sea demasiado arriesgado asignar una tarea importante a un robot que no tenga algún tipo de "introspección" sobre sus propias habilidades. Si no queremos que empiece proyectos que no vaya a acabar, será mejor hacerle saber qué es lo que es capaz de hacer. Si queremos hacerlo lo suficientemente versátil como para resolver nuevos tipos de problemas, tendrá que ser capaz de entender cómo resuelve los problemas más simples. En otras palabras, es bastante probable que un programa dedicado a resolver problemas realmente robusto necesite entenderse lo suficiente como para cambiarse. Si seguimos así, ¿por qué no van a llegar estas criaturas artificiales a alcanzar una vida mental mucho más rica que la de las personas? La evolución ha limitado el cableado de nuestros cerebros de muchas formas. Pero aquí tenemos muchas más opciones, podemos cablear nuestras máquinas como deseemos.

Pasará bastante tiempo antes de que aprendamos lo suficiente sobre el sentido común y su forma de razonar como para hacer máquinas tan inteligentes como el ser humano. Hoy en día ya sabemos bastante sobre cómo hacer sistemas "expertos" útiles y altamente especializados. Aún no sabemos como hacer que sean capaces de mejorarse de forma eficiente. Pero cuando respondamos a estas preguntas, tendremos que enfrentarnos a una mucho más extraña. Cuando aprendamos a hacerlo, ¿deberíamos construir máquinas en cierto sentido "mejores" que nosotros? Tenemos suerte de dejar esta pregunta a las generaciones futuras. Estoy seguro de que no querrán construir este tipo de cosas a no ser que tengan una buena razón para ello.

Tal y como la evolución cambió nuestra la forma de ver la vida, la IA cambiará la visión sobre la mente. A medida que encontremos más caminos para que las máquinas se comporten de un modo más sensible, aprenderemos más sobre nuestros procesos mentales. Recorriendo este sendero, encontraremos nuevas formas de considerar el "pensamiento" y los "sentimientos". Nuestra forma de verlos cambiará, haciendo que lo que ahora son inexpugnables misterios se conviertan en complejas –pero al tiempo comprensibles– redes de métodos para la representación y uso de las ideas. Esas ideas, a su vez, llevarán a la construcción de nuevas máquinas, lo que a su vez nos dará nuevas ideas. Nadie puede decirnos donde nos llevará todo esto y lo único que es seguro es que hoy en día, es absurdo hablar de cualquier tipo de diferencia básica entre la mente del ser humano y la de una hipotética máquina futura.