

La Revolución de la IA: El camino a la Súper Inteligencia

Estamos al borde de un cambio comparable a la creación de la vida humana en la Tierra. ¿Qué se siente estar parados aquí?

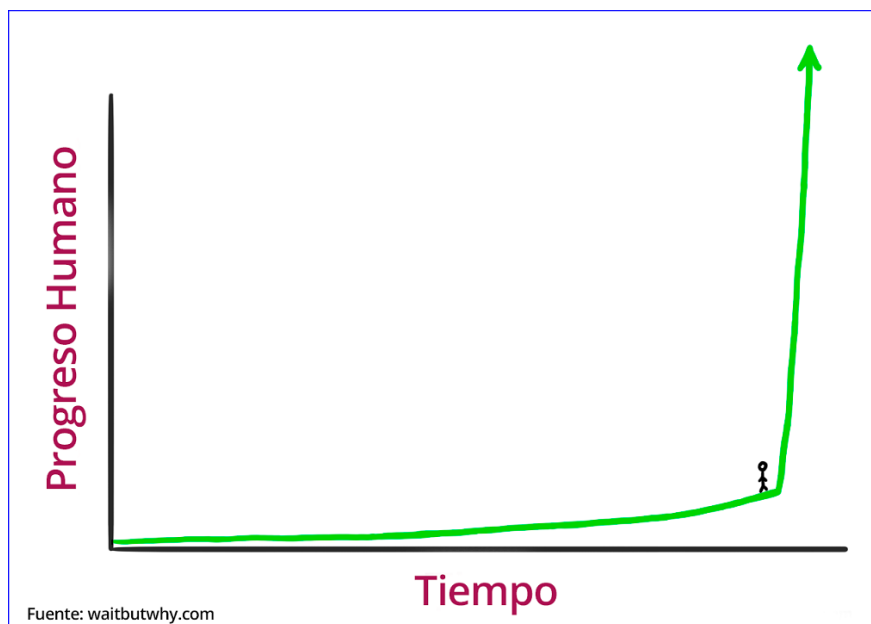
Por Tim Urban

Wait but Why?

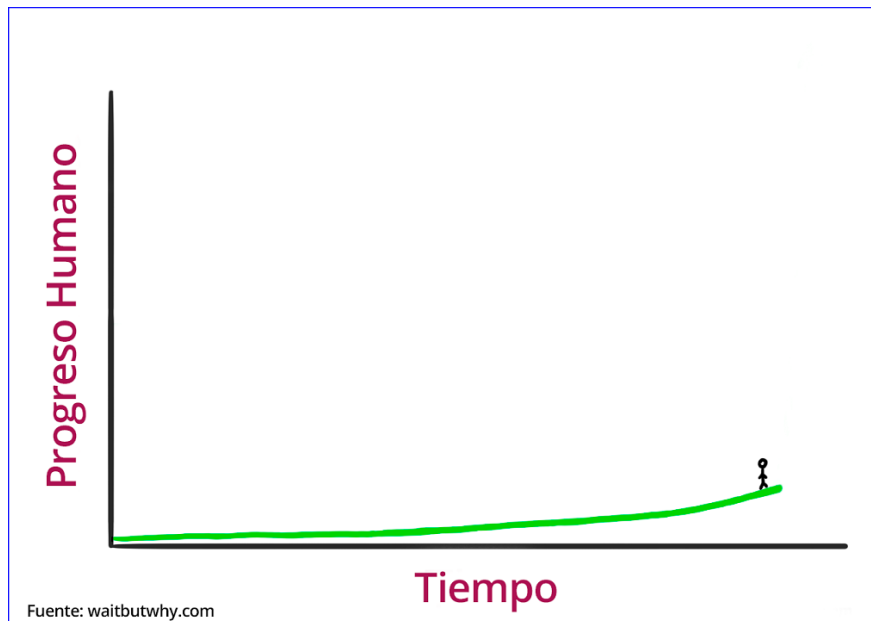
22 de enero de 2015

Traducido y glosado por *Lampadía*

(Ver resumen de este trabajo en: [Una explicación integral sobre el desarrollo de la Inteligencia Artificial](#)).



Parece ser un lugar bastante intenso para estar parado - pero tenemos que recordar algo sobre qué se siente al estar parado en un gráfico de tiempo: no se puede ver lo que está a la derecha. Así que así es como realmente se siente estar parado allí:



Lo que probablemente se siente bastante normal...

El futuro lejano – Llegará pronto

Imaginemos que tenemos una máquina del tiempo que nos lleve a 1750, una época en que el mundo tenía un escasez de energía permanente, la comunicación a larga distancia se limitaba a gritar en voz alta o disparar un cañón al aire, y todo el transporte se alimentaba de heno. Imaginemos que si al llegar a esa época, tomamos a un amigo, lo traemos al 2015, y luego lo llevamos a caminar por las calles y ver cómo reacciona a todo. Es imposible para nosotros entender cómo sería para él ver cápsulas brillantes moverse sobre una carretera, hablar con personas que se encuentran al otro lado del océano instantáneamente, ver deportes que están sucediendo a 1,000 millas de distancia, escuchar una composición musical que sucedió hace 50 años, y jugar con mi 'rectángulo mágico' que es capaz de capturar una imagen real o grabar una situación, generar un mapa en movimiento con un punto azul que muestra dónde se encuentra, mirar a la cara a alguien que se encuentra al otro lado del país, y otras increíbles magias en nuestro mundo. Todo esto es antes de mostrarle el Internet o explicarle cosas como la Estación Espacial Internacional, el gran colisionador Hadron, las armas nucleares, o la relatividad general.

Esta experiencia no sería solo sorprendente o chocante o incluso alucinante- esas palabras no son lo suficientemente grandes. En realidad podría morir.

Y entonces ¿qué pasaría si, después de morir, él se pusiera celoso y quisiera hacer lo mismo? Si regresara 12,000 años al pasado, al 24,000 ACy toma a un joven y lo lleva al 12,000 AC, le mostrara todo y el hombre respondiera diciendo, "Bueno, ¿cuál es tu punto? ¿A quién le importa?" Para que el individuo de 12,000 AC tenga la misma reacción, tendría que regresar en el tiempo más de 100,000 años y conseguir a alguien que pudiera mostrarle el fuego y el lenguaje por primera vez.

Para que alguien sea transportado al futuro y muera por el nivel de desconcierto, tiene que viajar bastantes años hasta llegar a un "nivel de progreso de muerte", o un Progreso de Muerte Unitario (PMU). Por lo que una DMU tomó 100,000 años en la época de cazadores-recolectores, pero posteriormente a la Revolución Agrícola, sólo tarda unos 12,000 años. El mundo posterior a la Revolución Industrial se ha movido tan rápidamente que una persona sólo necesita ir 1,750 años hacia adelante para lograr una DMU.

Este patrón (el progreso humano se mueve más y más rápido conforme pasa el tiempo) es lo que el futurista Ray Kurzweil llama **la ley de Rendimientos Acelerados de la historia humana**. Esto sucede porque las sociedades más avanzadas tienen la capacidad para avanzar a un ritmo más rápido que las sociedades menos avanzadas, porque son más avanzadas.

Esto también funciona en escalas más pequeñas. La película Volver al Futuro salió en 1985, y "el pasado" sucedió en el año 1955. En la película, cuando Michael J. Fox volvió a 1955, se sorprendió por la novedad de los televisores, los precios de las gaseosas, la falta de amor por la estridente guitarra eléctrica, y los cambios en las jergas. Era un mundo diferente, sí, pero si la película se hiciera hoy y el pasado se llevara a cabo en 1985, la película podría haber sido mucho más divertida con diferencias mucho más grandes. El personaje sería de un mundo antes de las computadoras, Internet y teléfonos móviles- y el Marty McFly de ahora, un adolescente nacido a finales de los años 90, estaría mucho más fuera de lugar en el año 1985 que el Marty McFly de la película en 1955.

Esto es por la misma razón que acabamos de discutir, la ley de Rendimientos Acelerados. La tasa media de avance entre 1985 y 2015 fue más alta que la tasa entre 1955 y 1985, ya que el primero fue un mundo más avanzado; **han ocurrido muchos más cambios en los últimos 30 años que en los 30 anteriores.**

Por lo tanto, los avances son cada vez más y más grandes y pasan más y más rápidamente. Esto sugiere algunas cosas muy intensas sobre nuestro futuro, ¿verdad?

Kurzweil sugiere que el progreso total de todo el siglo 20 se habría logrado en tan sólo 20 años a un ritmo de avance del año 2000. En otras palabras, al año 2000, la tasa de progreso fue cinco veces más rápida que la tasa media de progreso en el siglo 20. Además, Kurzweil cree que el valor del avance de otro siglo 20 pasó tan solo entre 2000 y 2014 y que el valor de progreso de otro siglo 20 va a pasar en el año 2021, en sólo siete años. En un par de décadas, cree que el valor del avance de otro siglo 20 va a pasar varias veces en el mismo año, e incluso más tarde, en menos de un mes. Considerando todo, a causa de la Ley de Rendimientos Acelerados, **Kurzweil cree que el siglo 21 tendrá 1,000 veces el progreso de del siglo 20.**

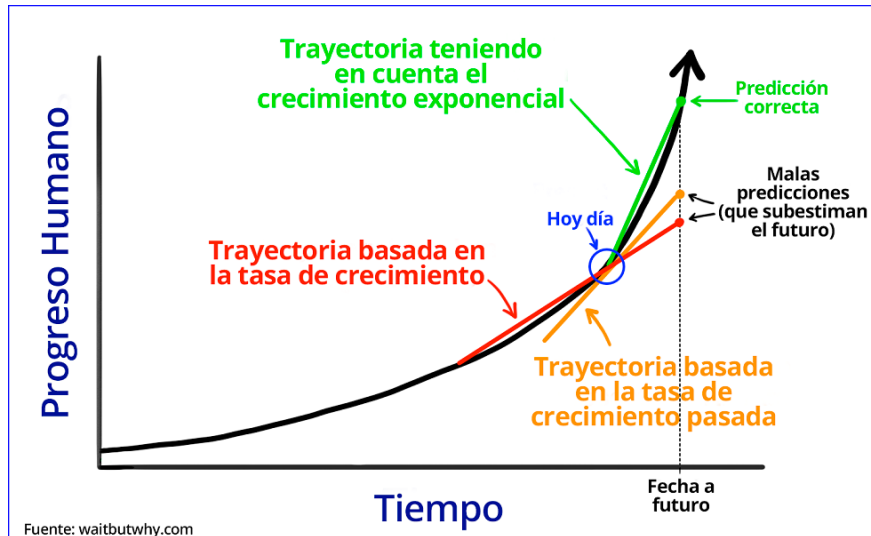
Si es que Kurzweil (y otros que están de acuerdo con él) están en lo correcto, entonces podemos estar tan impresionados en 2030 como nuestro hombre de 1750 lo estaría en 2015, es decir, la próxima DMU sólo tomará un par de décadas, y el mundo en 2050 podría ser tan diferente que los del mundo de hoy apenas lo reconoceremos.

Esto no es ciencia ficción. Es lo que muchos científicos más inteligentes y con más conocimiento que nosotros creen firmemente y, si nos fijamos en la historia, es lo que debemos predecir lógicamente.

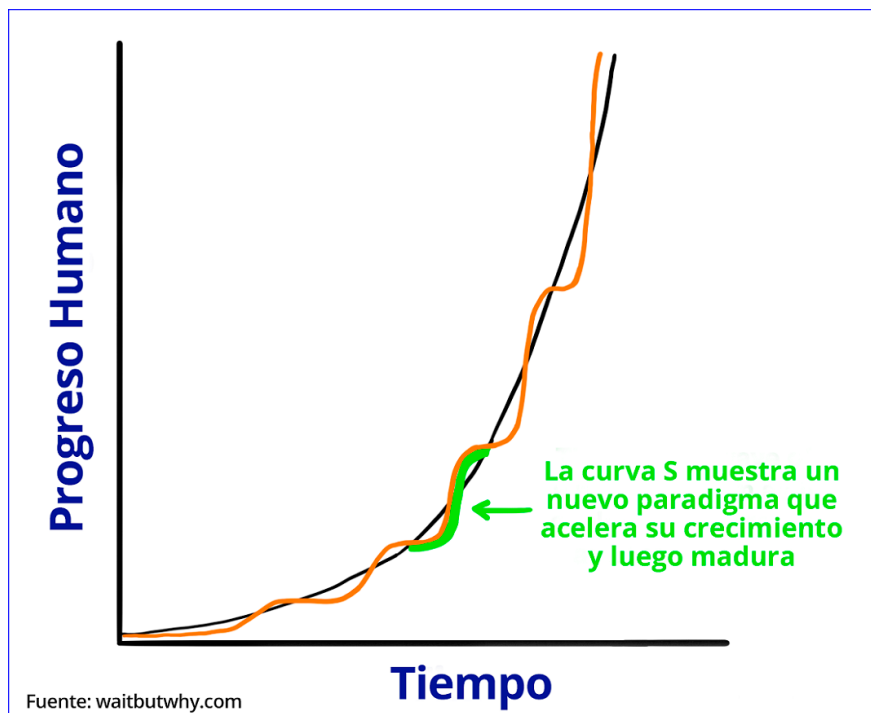
Así que ¿por qué, cuando me escuchen decir algo como "dentro de 35 años el mundo podría ser totalmente irreconocible," pensamos, "Chévere... pero noooo"? Tres razones por las que somos escépticos de las predicciones extravagantes del futuro:

- 1) **Cuando se trata de la historia, pensamos en líneas rectas.** Cuando imaginamos el progreso de los próximos 30 años, miramos hacia atrás, hacia el progreso de los 30 años anteriores como un indicador de cuánto es probable que suceda. Cuando pensamos en cuánto cambiará el mundo en el siglo 21, sólo tomamos el progreso del siglo 20 y lo añadimos al año 2000. Es más intuitivo para nosotros **pensar linealmente**, cuando deberíamos **pensar exponencialmente**. **Para**

pensar correctamente en el futuro, es necesario imaginar que las cosas que se mueven a un ritmo mucho más rápido de lo que están haciendo ahora.



2. La trayectoria de la historia más reciente a menudo cuenta una historia distorsionada. En primer lugar, incluso una empinada curva exponencial parece lineal cuando sólo se mira a una pequeña porción de la misma, de la misma manera que si nos fijamos en un pequeño segmento de un círculo, se ve casi como una línea recta. En segundo lugar el crecimiento exponencial no es totalmente liso y uniforme. Kurzweil explica que el progreso ocurre en "curvas S":



Una S es creada por la ola de progreso cuando un nuevo paradigma se extiende por el mundo. La curva pasa por tres fases:

1. El crecimiento lento (la primera fase de crecimiento exponencial)
2. El crecimiento rápido (el tardío y explosivo crecimiento exponencial)
3. El nivelamiento cuando un paradigma en particular madura

Si nos fijamos solamente en la historia más reciente, la parte de la curva S en la que nos encontramos en este momento puede oscurecer nuestra percepción de cuán rápido están avanzando las cosas. El periodo de tiempo entre 1995 y 2007 en el que se produjo la explosión de la Internet, vio la creación de Microsoft, Google y Facebook en la conciencia pública, el nacimiento de las redes sociales y la introducción de teléfonos celulares y teléfonos inteligentes. Esa fue la Fase 2: el crecimiento de la S.

Sin embargo, el periodo de 2008 hasta 2015 ha sido menos innovador, al menos en el frente tecnológico. Si alguien pensara en el futuro hoy, podría examinar los últimos años para medir el ritmo actual de avance, pero no estaría viendo la realidad completa. De hecho, un nuevo y enorme crecimiento de Fase 2 podría estar gestándose en estos momentos.

3) Nuestra propia experiencia nos convierte en hombres viejos y obstinados sobre el futuro. Basamos nuestras ideas sobre el mundo en nuestra experiencia personal y esa experiencia ha arraigado en nuestras cabezas que la tasa de crecimiento de los últimos años es "la forma como suceden las cosas." También estamos limitados por nuestra imaginación, que toma nuestra experiencia y la utiliza para evocar las predicciones futuras, pero a menudo, lo que sabemos simplemente no nos da las herramientas para pensar con precisión acerca del futuro. Cuando oímos que una predicción sobre el futuro contradice nuestra experiencia basada en la noción de cómo funcionan las cosas, nuestro instinto es que la predicción debe ser ingenua.

Si digo, más adelante en este post, que podremos vivir hasta los 150 o 250 años, o ser inmortales, nuestro instinto será: "Eso es estúpido, si hay una cosa que sé de la historia, es que todo el mundo muere." Y sí, nadie en el pasado no ha muerto. Pero nadie voló aviones antes de que se inventaran.

El hecho es que, si estamos siendo verdaderamente lógicos y esperando que los patrones históricos continúen, debemos concluir que mucho, mucho, mucho más debe cambiar en las próximas décadas de lo que intuitivamente esperamos.

El camino a la súperinteligencia

¿Qué es IA?

Si son como yo, solías pensar que la Inteligencia Artificial era un tonto concepto de ciencia ficción, pero últimamente se le menciona por la gente seria y no entendemos muy bien por qué.

Hay tres razones que mucha gente no entiende bien el término AI:

- 1) **La IA se asocia con películas.** Guerra de las Galaxias. Terminator. 2001: Una odisea del espacio. Incluso los Supersónicos. Y son de género ficción, al igual que los personajes robots. Así que la IA nos tiene que sonar un poco a ficción.
- 2) **La IA es un tema muy amplio.** Se extiende desde la calculadora del celular a vehículos de auto-conducción a algo que podría cambiar el mundo de manera espectacular en el futuro. La IA se refiere a todas estas cosas, lo cual es confuso.
- 3) **Utilizamos la IA todo el tiempo en nuestra vida diaria, pero a menudo no nos damos cuenta de que es IA.** John McCarthy, quien acuñó el término "inteligencia artificial" en

1956, se quejó de que "tan pronto funcione, nadie lo volverá a llamar IA." Debido a este fenómeno, la IA a menudo suena más como una predicción mítica del futuro que de una realidad.

Así que vamos a aclarar las cosas. En primer lugar, dejemos de pensar en robots. Un robot es un contenedor de IA, a veces imitando la forma humana, a veces no, pero la IA es el equipo en el interior del robot. El IA es el cerebro, y el robot es su cuerpo, si es que tiene un cuerpo. Por ejemplo, el software y los datos detrás de Siri es IA, la voz de la mujer que oímos es una personificación de la IA y no hay ningún robot involucrado en lo absoluto.

En segundo lugar, probablemente hemos escuchado el término "singularidad" o "singularidad tecnológica". Este término se ha utilizado en las matemáticas para describir una situación similar a la asíntota, donde las reglas normales ya no se aplican. Ha sido utilizado en la física para describir un fenómeno como un agujero negro infinitamente pequeño, denso o el punto en el que todos estábamos apiñados justo antes del Big Bang. Una vez más, situaciones en las que no se aplican las reglas habituales. En 1993, Vernor Vinge escribió un famoso ensayo en el que se aplica el término para el momento en el futuro cuando la inteligencia de nuestra tecnología sea superior a nuestra -un momento en el que la vida tal como la conocemos cambiará para siempre y ya no se aplicarán las reglas normales. Entonces Ray Kurzweil confundió un poco las cosas mediante la definición de la singularidad como el momento en que la ley de rendimientos acelerados llega a un ritmo tan extremo que el progreso tecnológico ocurriría a un ritmo aparentemente infinito y después del cual vamos a estar viviendo un nuevo mundo.

Por último, si bien hay muchos tipos o formas de IA diferentes, las categorías fundamentales que necesitamos se basan en el calibre de una IA. Hay tres categorías principales de calibre de IA:

Calibre 1) Una Inteligencia Artificial estrecha (ANI, por sus siglas en inglés: Artificial Narrow Intelligence): A veces se denomina IA Débil. La Inteligencia Artificial Estrecha se especializa en un área. Hay IA que puede vencer al campeón mundial de ajedrez en ajedrez, pero eso es lo único que lo hace. Si le preguntas cuál es la mejor manera de almacenar los datos en un disco duro, te mirará sin comprender.

Calibre 2) Una Inteligencia Artificial General (AGI, por sus siglas en inglés: Artificial General Intelligence): Algunas veces conocido como la IA fuerte o la IA a nivel humano. La Inteligencia Artificial General se refiere a un equipo que es tan inteligente como un ser humano en todos los ámbitos, una máquina que puede realizar cualquier tarea intelectual que un ser humano pueda hacer. La profesora Linda Gottfredson describe la inteligencia como "una capacidad mental muy general que, entre otras cosas, implica la capacidad de razonar, planear, resolver problemas, pensar de manera abstracta, comprender ideas complejas, aprender rápidamente y aprender de la experiencia." AGI sería capaz de hacer todas esas cosas tan fácilmente como sea posible.

Calibre 3) Una Superinteligencia Artificial (por sus siglas en inglés: Artificial Super Intelligence): Un filósofo y líder pensador de IA de Oxford. Nick Bostrom, define la super-inteligencia como "un intelecto que es mucho más inteligente que los mejores cerebros humanos en prácticamente todos los campos, incluyendo la creación científica, la sabiduría general y las habilidades sociales." La superinteligencia varía de un equipo que es sólo un poco más inteligente que un ser humano a otro que es miles de millones de veces más inteligente en todos los ámbitos.

A partir de ahora, los seres humanos han conquistado el calibre más bajo de la IA, de muchas maneras y está en todas partes. La Revolución de la IA es el camino de la ANI, a través de IAG,

hacia SIA- un camino que puede o no puede sobrevivir, pero que, en cualquier caso, va a cambiar todo.

Vamos a echarle un vistazo a lo cómo creen los principales pensadores en el campo que se verá este camino y por qué esta revolución puede ocurrir más pronto de lo que parece:

Dónde Estamos Actualmente, un mundo que se ejecuta en ANI

La Inteligencia Artificial estrecha es la inteligencia artificial que es igual o más inteligente que un humano en una cosa específica. Unos cuantos ejemplos:

- Los autos están llenos de sistemas ANI
- El teléfono es una pequeña fábrica de ANI
- El filtro de spam del correo electrónico es un tipo clásico de ANI
- Esa cosa espeluznante que se aparece cuando buscas un producto en Amazon y luego el mismo producto aparece en "recomendaciones para usted" en una web diferente, o cuando Facebook de alguna manera sabe a quiénes recomendar como amigos. Eso es una red de sistemas ANI
- Google Translate es otro sistema clásico de ANI
- Los mejores jugadores de Damas, Ajedrez, Scrabble, Backgammon, y Othello del mundo son ahora sistemas de ANI
- La búsqueda de Google es un gran cerebro ANI con métodos muy sofisticados para la clasificación de páginas.
- Y esos son sólo en el mundo de los consumidores. Los sofisticados sistemas ANI son ampliamente utilizados en los sectores e industrias militares, de manufactura y finanzas (los operadores de algoritmos de alta frecuencia de IA representan más de la mitad de las acciones de capital negociados en los mercados americanos), y en los sistemas de expertos, como los que ayudan a los médicos a hacer diagnósticos y, lo más famoso, el IBM Watson, que contenía suficientes elementos de juicio y entendió suficientemente bien a Trebek hablar como para vencer a los campeones más experimentados de Jeopardy

El camino de la ANI A IAG

Por qué es tan difícil

Nada nos hará apreciar tanto la inteligencia humana como aprender acerca de lo increíblemente difícil que es tratar de crear una computadora tan inteligente como nosotros.

Lo que es interesante es que las partes difíciles de tratar de construir una AGI (un equipo tan inteligente como los humanos en general, no sólo en una especialidad) no son intuitivamente los que uno pensaría que sean. Construir un equipo que pueda multiplicar dos números de diez dígitos en una fracción de segundo es increíblemente fácil. Construir una que pueda mirar a un perro y responder si es un perro o un gato- espectacularmente difícil. ¿Hacer que una IA pueda vencer a cualquier ser humano en el ajedrez? Hecho. ¿Hacer uno que pueda leer un párrafo de un libro de imágenes de seis años de edad y no sólo reconocer las palabras, sino entender su significado? Google está gastando miles de millones de dólares tratando de hacerlo. Cosas difíciles, como el cálculo, las estrategias del mercado financiero y la traducción del lenguaje, son abrumadoramente fáciles para una computadora, mientras que las cosas fáciles como la visión, el movimiento y la percepción son terriblemente difíciles para una IA. O, como lo dijo un

científico de la computación, Donald Knuth, "La IA ya ha tenido éxito en hacer básicamente todo lo que requiere 'pensar', pero ha dejado de hacer la mayor parte de lo que la gente y los animales hacen 'sin pensar'."

Lo que uno se da cuenta rápidamente cuando analiza esto es que esas cosas que parecen fáciles para nosotros en realidad son increíblemente complicadas, y sólo parece fácil porque esas habilidades se han optimizado en nosotros (y en la mayoría de los animales) durante cientos de millones de años de evolución animal. Al mover una mano hacia un objeto, los músculos, los tendones y los huesos del hombro, el codo y la muñeca realizan al instante una larga serie de operaciones físicas, en conjunto con los ojos, que le permiten a la persona mover la mano en una línea recta a través de tres dimensiones. Parece como si no causara mucho esfuerzo para la persona, porque ésta ha perfeccionado el software en su cerebro para hacerlo. La misma idea se aplica a por qué no es que el malware sea tonto por no ser capaz de descifrar la letra cursiva, sino es que el cerebro es muy impresionante por ser capaz de hacerlo.

Por otra parte, la multiplicación de números grandes o jugar al ajedrez son nuevas actividades para las criaturas biológicas y no hemos tenido tiempo para desarrollar esa habilidad, por lo que una computadora no tiene que trabajar mucho para lograrlo. Analicémoslo. ¿Qué preferiríamos lograr, construir un programa que pueda multiplicar números grandes o uno que pueda comprender la esencia de una letra 'B' tan bien que podría descifrar una B en cualquiera de los miles de impredecibles formatos de escritura a mano y pueda saber inmediatamente que es una B?

Para ser de nivel humano inteligente, una computadora tendría que entender cosas como la diferencia entre las sutiles expresiones faciales, la distinción entre estar satisfecho, aliviado, contento, campante y jubiloso.

Entonces, ¿cómo llegamos?

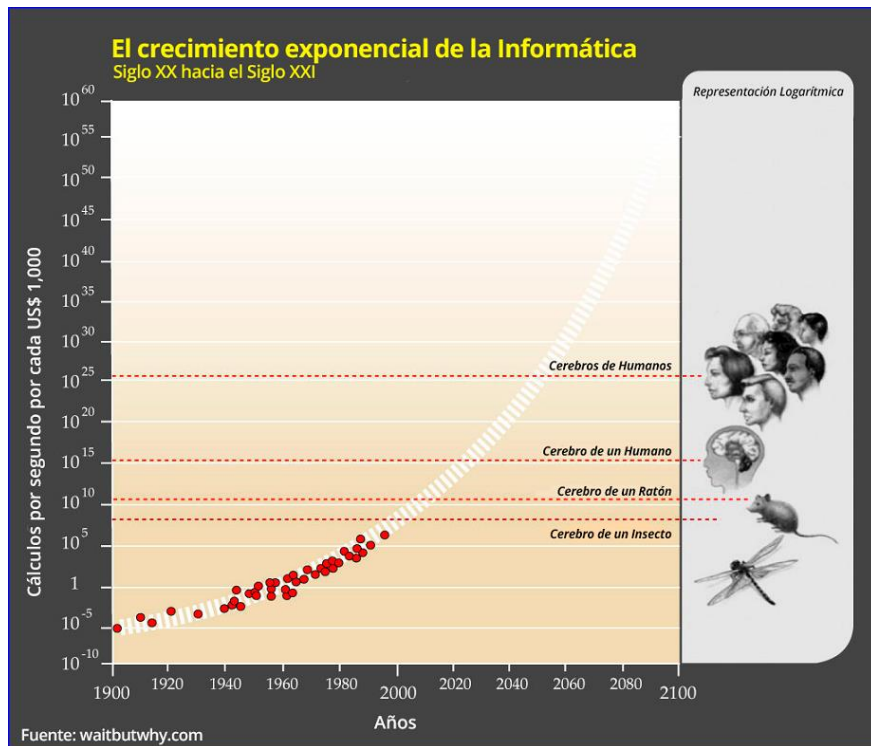
La primera clave para crear AGI: El aumento en la potencia de cálculo

Una cosa que sin duda tiene que pasar para que la AGI sea una posibilidad es un aumento en la potencia del hardware de la computadora. Si un sistema de inteligencia artificial va a ser tan inteligente como el cerebro, tendrá que ser igual a la capacidad bruta de la computadora del cerebro.

En la actualidad, el superordenador más rápido del mundo, la Tianhe-2 de China, en realidad ha superado esa capacidad, llegando a 34 billones de billones de cps. Pero Tianhe-2 también es un espeso, ocupando 720 metros cuadrados, con 24 megavatios de potencia (el cerebro funciona con sólo 20 watts), y un costo de US\$ 390 millones para construirlo. No es especialmente aplicable para la utilización masiva, o incluso para el uso comercial o industrial.

Kurzweil sugiere que pensamos en el estado de computadoras analizando cuántos cps se pueden comprar por US\$ 1,000. Cuando ese número alcance el nivel humano-10 cuatrillones de cps-entonces significará que el AGI podría llegar a ser una parte muy real de la vida.

La Ley de Moore es una regla históricamente fiable que afirma que el potencial máximo computacional del mundo se duplicará aproximadamente cada dos años, lo que significa que el avance del hardware del equipo, al igual que el progreso humano a través de la historia, crecerá exponencialmente. En cuanto a cómo esto se relaciona con la idea de Kurzweil de cps/ US\$ 1,000, estamos actualmente en alrededor de 10 billones de cps / US\$ 1,000, al mismo ritmo de trayectoria prevista en este gráfico:



Así que las computadoras de US\$ 1,000 del mundo ahora le ganan al cerebro de un ratón y están a aproximadamente una milésima del nivel humano. Esto no suena como mucho hasta que recordamos que estábamos en alrededor de una billonésima del nivel humano en 1985, una mil millonésima en 1995 y una millonésima en 2005. Al estar en una milésima en 2015, nos indica que llegaremos a tener una computadora que compita con el poder del cerebro a un precio asequible para el año 2025.

Así que en el lado del hardware, el poder primario necesario para AGI está técnicamente disponible ahora.

Segunda clave para crear AGI: Volverlo inteligente

Esta es la parte complicada. La verdad es que nadie sabe realmente cómo hacer que sea inteligente- estamos todavía debatiendo cómo hacer que un equipo de inteligencia de nivel humano sea capaz de saber lo que es un perro y una 'B' escrita en cursiva e identificar una película mediocre. Pero hay un montón de estrategias inverosímiles y, en algún momento, una de ellas va a funcionar. Aquí están las tres estrategias más comunes que encontré:

1) Plagiar el cerebro.

El mundo de la ciencia está trabajando duro en la ingeniería inversa del cerebro para averiguar cómo la evolución hizo algo tan impresionante, las estimaciones optimistas dicen que podremos hacerlo en 2030. Una vez que lo hagamos, vamos a saber todos los secretos de cómo funciona el cerebro y podremos inspirarnos de ella y robar sus innovaciones. Un ejemplo de la arquitectura del ordenador que imita al cerebro es la red neuronal artificial. Comienza como una red de "neuronas transistores", conectadas entre sí con entradas y salidas, y que no sabe nada, parecido a un cerebro infantil. La forma en que se "aprende" es que trata de hacer una tarea, por ejemplo el reconocimiento de escritura, y en un primer momento, sus disparos neuronales y posteriores conjeturas en descifrar cada letra serán completamente al azar. Pero cuando hizo

algo bien, las conexiones de transistores para crear esa respuesta se fortalecen; cuando se le afirma que es un error, las conexiones de esas vías se debilitan. Después de una gran cantidad de prueba y error, la red ha formado, por sí misma, vías neuronales inteligentes y la máquina se ha optimizado para la tarea. El cerebro aprende más o menos de esta manera, pero de una manera más sofisticada, y a medida que seguimos estudiando el cerebro, estamos descubriendo nuevas formas ingeniosas para aprovechar los circuitos neuronales.

El plagio más extremo implica una estrategia denominada "la emulación de todo el cerebro", donde el objetivo es cortar un cerebro real en capas delgadas, escanear cada una, utilizar un software para reconstruir un modelo 3-D exacto, y luego implementar el modelo en una computadora potente.

¿Cuán lejos estamos de alcanzar la emulación de todo el cerebro? Bien, hasta ahora, recientemente hemos sido capaces de emular un cerebro de un gusano de 1 mm de largo, que se compone de sólo 302 neuronas totales. El cerebro humano contiene 100 mil millones. Si eso lo hace parecer como un proyecto sin esperanza, recuerda el poder del progreso exponencial-ahora que hemos conquistado el diminuto cerebro de un gusano, podremos lograr la de una hormiga en poco tiempo, seguido de un ratón y, de repente, esto parecerá mucho más plausible.

2) Tratar de que la evolución haga lo que hizo antes, pero para nosotros esta vez.

Así que si decidimos que el examen del niño inteligente es demasiado difícil como para copiarlo, podemos tratar de copiar la forma en que estudia para las pruebas.

Aquí hay algo que sabemos. La construcción de un equipo tan potente como el cerebro es posible, la evolución de nuestro propio cerebro es la prueba. Y si el cerebro es demasiado complejo para que nosotros lo emulemos, podríamos intentar emular su revolución en su lugar. El hecho es que, incluso si podemos emular un cerebro, que podría ser como intentar construir un avión copiando los movimientos de aleteo-frecuencia de un pájaro, las máquinas están mejor diseñadas utilizando un nuevo enfoque, orientado a máquinas, no mediante la imitación exacta de la biología.

Entonces, ¿cómo podemos simular la evolución para construir un AGI? El método, llamado "algoritmos genéticos", funcionaría así: habría un proceso de rendimiento y evaluación que funcionaría una y otra vez (de la misma manera que las criaturas biológicas "rinden" para vivir la vida y es "evaluada" por el hecho de si las arreglan para reproducirse o no). Un grupo de equipos trataría de hacer las mismas tareas, y los más exitosos serían reproducidos entre sí incorporando la mitad de cada uno de sus programas en un nuevo equipo. Las menos exitosas serían eliminadas. Durante muchas, muchas iteraciones, este proceso de selección natural produciría cada vez mejores ordenadores. El reto sería la creación de un ciclo de evaluación y reproducción automatizada para que este proceso de evolución funcione por sí solo.

La desventaja de copiar la evolución es que la evolución toma unos mil millones de años en hacer las cosas y queremos hacer esto en unas pocas décadas.

Pero le llevamos muchas ventajas a la evolución. En primer lugar, la evolución no prevee y funciona de forma aleatoria- produciendo más mutaciones inútiles que útiles, pero nosotros podemos controlar el proceso para que sólo impulse mutaciones beneficiosas y específicos. En segundo lugar, la evolución no tiene un objetivo específico, como mayor inteligencia, a veces un ambiente incluso podría seleccionar tener menos inteligencia superior (ya que utiliza una gran cantidad de energía). Nosotros, por el otro lado, podríamos dirigir específicamente este proceso evolutivo hacia el aumento de la inteligencia. En tercer lugar, para seleccionar la inteligencia, la

evolución tiene que innovar en muchos otros ámbitos para facilitar la inteligencia (como la modernización de las maneras células que producen energía), mientras que nosotros podemos eliminar esas cargas adicionales y usar cosas como la electricidad. No hay duda de que seríamos mucho, mucho más rápidos que la evolución, pero aún no está claro si vamos a ser capaces de mejorar la evolución lo suficiente como para hacerla una estrategia viable.

3) Volver todo esto un problema para la computadora, no para nosotros.

Esto es cuando los científicos se desesperan y tratan de programar la prueba para que se la tome a sí misma. Pero podría ser el método más prometedor que tenemos.

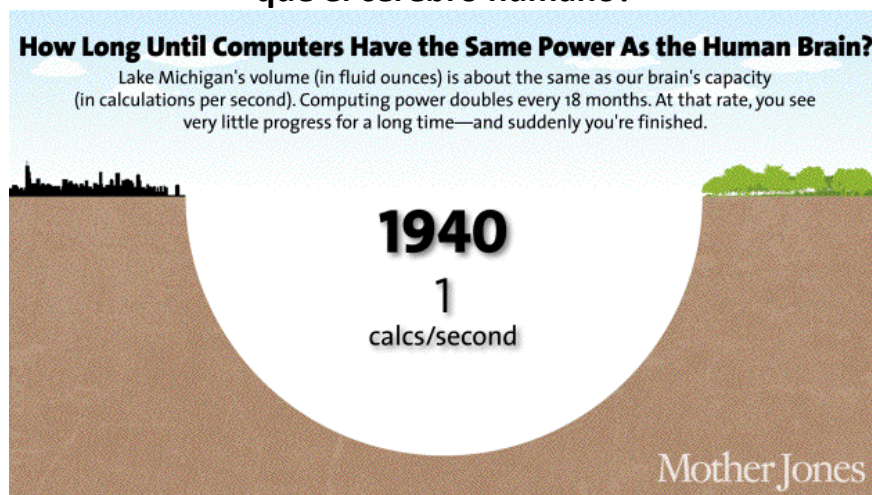
La idea es que construiríamos una computadora cuyas dos principales habilidades serían investigar sobre la IA y cambiar códigos en sí mismo, lo que le permitiría no sólo aprender sino mejorar su propia arquitectura. Le enseñaríamos a las computadoras a ser científicos informáticos para que puedan desplegar su propio desarrollo. Y ese sería su principal trabajo: encontrar la manera aumentar su inteligencia. Más sobre esto más adelante.

Todo esto podría ocurrir pronto

Los rápidos avances en el hardware y la experimentación innovadora con el software están ocurriendo simultáneamente, y la AGI podría llegar más rápido e inesperadamente por dos razones principales:

- 1) El crecimiento exponencial es intenso y lo que parece ser un ritmo muy lento de avance puede crecer rápido repentinamente, este GIF ilustra este concepto muy bien:

¿Cuánto tiempo hasta que las computadoras tengan la misma potencia que el cerebro humano?



El volumen del lago Michigan (en onzas de fluido) es aproximadamente el mismo que la capacidad de nuestro cerebro (en cálculos por segundo). la potencia de computación se duplica cada 18 meses. A ese ritmo, se ve muy poco progreso durante mucho tiempo- y, de pronto, todo se terminó.

(Para ver la gráfica interactiva, ver en [Lampadia: Una explicación integral sobre el desarrollo de la Inteligencia Artificial](#))

2) Cuando se trata de software, el progreso puede parecer lento, pero luego una epifanía puede cambiar la velocidad de avance (algo así como en la ciencia, cuando los humanos pensaban que el universo era geocéntrico, tenían dificultades para calcular cómo funcionaba, pero luego del descubrimiento de que era heliocéntrico, de repente todo fue mucho más fácil). O bien, cuando se trata de una computadora que mejora por sí misma, puede parecer muy lejano, pero en realidad es sólo un pequeño cambio en el sistema que puede volverse 1,000 veces más eficaz y exponencial para la inteligencia de nivel humano.

El camino de la AGI hacia la ASI

En algún momento, habremos logrado que el AGI- computadoras con inteligencia a nivel humano. Sólo un grupo de personas y equipos conviviendo en igualdad.

Cuando el AGI tenga el mismo nivel de inteligencia y capacidad computacional de un humano, tendría ventajas significativas sobre los humanos. Por ejemplo:

Hardware:

1. **Velocidad.** Las neuronas del cerebro se maximizan alrededor de 200 Hz, mientras que los microprocesadores de hoy (que son mucho más lentos que cuando llegue el AGI) llegan a 2 GHz, o 10 millones de veces más rápido que nuestras neuronas. Y las comunicaciones internas del cerebro, que pueden moverse a unos 120 m / s, son superadas por la capacidad de un equipo para comunicarse ópticamente a la velocidad de la luz.

2. **Tamaño y almacenamiento.** El cerebro es bloqueado en su tamaño por la forma de los cráneos, y que no podía ser mucho más grande de todos modos, las comunicaciones o 120 m/s internos sería demasiado largo para llegar de una estructura del cerebro a otra. Las computadoras pueden ampliarse a cualquier tamaño físico, lo que libera mucho hardware para utilizar, una mucho mayor memoria de trabajo (RAM) y una memoria a largo plazo (almacenamiento en disco duro) que tiene una mayor capacidad y precisión que la nuestra.

3. **La fiabilidad y la durabilidad.** No es sólo el recuerdo de un equipo que sería más preciso. Los transistores informáticos son más precisos que las neuronas biológicas y es menos probable que se deterioren (y pueden ser reparados o reemplazados si lo hacen). Los cerebros humanos también se fatigan con facilidad, mientras que las computadoras pueden correr sin parar, al máximo rendimiento, 24/7.

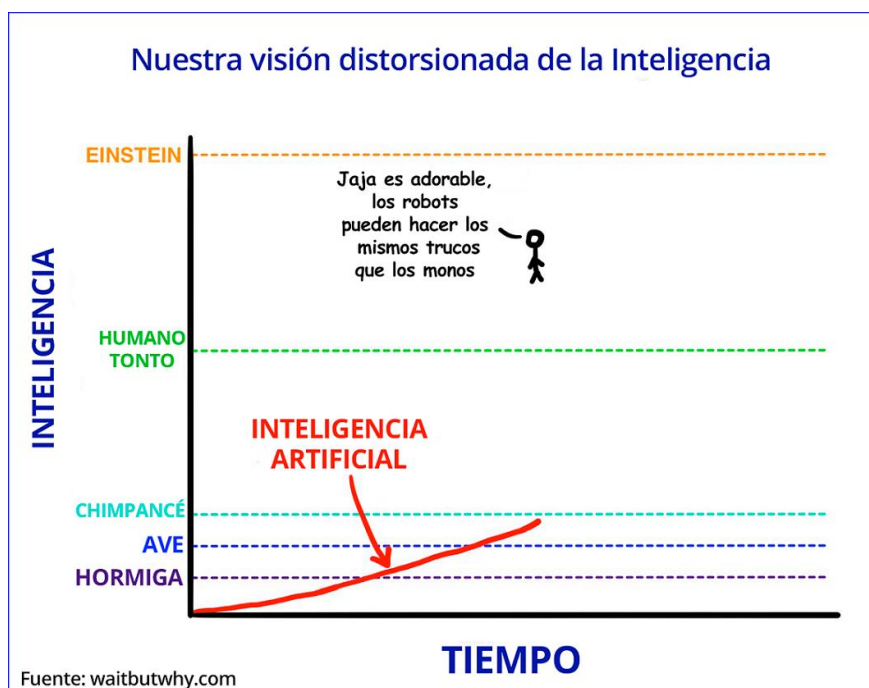
Software:

- **Editabilidad, capacidad de actualización, y una anchura más amplia de posibilidades.** A diferencia del cerebro humano, el software de computadora puede recibir actualizaciones y correcciones y puede ser fácilmente experimentado. Las mejoras también podrían extenderse a las zonas donde los cerebros humanos son débiles.
- **Capacidad colectiva.** Los seres humanos le ganan a todas las otras especies en la construcción de una gran inteligencia colectiva. Comenzando con el desarrollo del lenguaje y la formación de comunidades grandes y densas, avanzando a través de los inventos de escritura e impresión, y ahora a través de herramientas como el Internet, la inteligencia colectiva de la humanidad es una de las principales razones por las que hemos sido capaces de llegar por delante de todas las demás especies. Y las

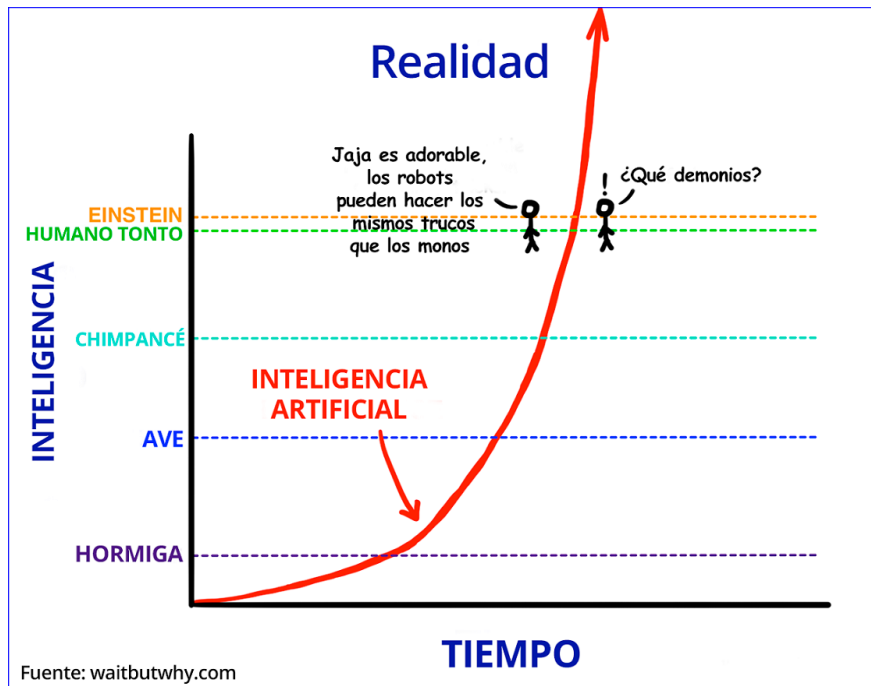
computadoras serán mejores que nosotros. Una red mundial de IA ejecutando un mismo programa podría sincronizar periódicamente todo lo que cualquier equipo aprenda al instante, actualizándose en todos los demás equipos. El grupo también podría tomar un objetivo como una unidad, porque no habrían opiniones disidentes y motivaciones e interés propio, como el que tenemos dentro de la población humana.

La IA, que probablemente llegará a AGI mediante la programación de auto-mejora, no vería la "inteligencia de nivel humano" como un hito importante- sólo es un hito relevante desde nuestro punto de vista- y no tendría ninguna razón para "parar" cuando llegue a nuestro nivel. Y teniendo en cuenta las ventajas que tendría la AGI sobre nosotros solo con una inteligencia equivalente a la humana, **es bastante obvio que sólo llegaría a la inteligencia humana por un breve instante antes de seguir avanzando rápidamente hacia una inteligencia superior a la humana.**

Puede que esto nos sorprenda muchísimos cuando suceda. La razón es que desde nuestra perspectiva, A), mientras que la inteligencia entre los diferentes tipos de animales varía, la principal característica es que somos conscientes que la inteligencia de cualquier animal es muy inferior a la nuestra, y B) vemos a los seres humanos más inteligentes como MUY por encima que los seres humanos más tontos. Algo así como esto:



Entonces, cuando la IA se acerque nuestro nivel de inteligencia, vamos a verla como simplemente se va volviendo más inteligente, para un animal. Luego, cuando llegue a la menor capacidad de los humanos -Nick Bostrom utiliza el término "el tonto del pueblo"- vamos a decir algo como, "Oh, wow, es como un ser humano tonto. ¡Genial!" La única cosa es que, en el gran espectro de la inteligencia, todos los seres humanos, desde el tonto del pueblo hasta Einstein, están dentro de un rango muy pequeño, por lo que poco después de alcanzar el nivel del tonto del pueblo y ser declarado AGI, de repente va a ser más inteligente que Einstein y no sabremos cómo paso:



Y, ¿qué sucede después de eso...?

Una explosión de Inteligencia

Espero que hayan disfrutado su tiempo normal, porque aquí es cuando este tema se vuelve anormal y da miedo, y así será de aquí en adelante. Quiero hacer una pausa aquí para recordarles que cada cosa que voy a decir es real —ciencia real y previsiones reales del futuro de una gran variedad de los pensadores más respetados y científicos. Hemos de tener que recordarlo.

De todos modos, como he dicho anteriormente, la mayoría de nuestros modelos actuales para llegar al AGI implican que la IA llegará por la auto-mejora. Y una vez que llegue a AGI, incluso los sistemas que se formaron y crecieron a través de métodos que no implicaron la auto-mejora ahora serían lo suficientemente inteligentes como para comenzar la auto-mejora si así lo quisieran.

Y aquí es donde llegamos a un concepto intenso: la auto-mejora recursiva. Funciona de la siguiente manera- Un sistema de inteligencia artificial a un cierto nivel, digamos un hombre tonto, se programa con el objetivo de mejorar su propia inteligencia. Una vez que lo hace, es más inteligente, tal vez en este momento es a nivel de Einstein. Ahora, cuando trabaja para mejorar su inteligencia, con una inteligencia del nivel de Einstein, se le hace mucho más fácil y se pueden hacer saltos más grandes. Estos saltos hacen que sea mucho más inteligente que cualquier ser humano, lo que le permite dar saltos aún más grandes. A medida que los saltos se hacen más grandes y ocurren con mayor rapidez, la AGI aumenta en inteligencia y pronto alcanza el nivel de superinteligencia de un sistema de ASI. Esto se llama una explosión de Inteligencia, que es el máximo ejemplo de la ley de rendimientos acelerados.

Existe cierto debate acerca de qué tan pronto la IA alcanzará un nivel humano de inteligencia. El año promedio dicho en una encuesta a cientos de científicos sobre cuando creyeron que sería probable que llegásemos fue 2040- eso es en sólo 25 años a partir ahora, lo

que no suena tan lejano hasta que se considera que muchos de los pensadores en este campo creen que es probable que la progresión de la AGI a la ASI sucede muy rápidamente.

Al igual que-esto podría suceder:

Se necesitan décadas para que el primer sistema de inteligencia artificial llegue a la inteligencia general de bajo nivel, pero finalmente sucederá. Una computadora es capaz de entender el mundo que lo rodea, así como un ser humano de cuatro años. De repente, una hora después de llegar a ese hito, el sistema bombea la gran teoría de la física que unifica la relatividad general y la mecánica cuántica, algo que ningún humano ha sido capaz de hacer definitivamente. 90 minutos después de eso, la IA se ha convertido en un ASI, 170,000 veces más inteligente que un ser humano.

Una super-inteligencia de esa magnitud no es algo que podemos captar de forma remota, no más que un abejorro puede entender la economía keynesiana. En nuestro mundo, la inteligencia significa un CI de 130 y la estupidez significa un CI de 85, aún no disponemos de una palabra para un CI de 12,952.

Lo que sí sabemos es que el dominio absoluto de los seres humanos en la Tierra sugiere una regla clara: **con la inteligencia viene el poder**. Lo que significa una ASI, cuando la creemos, será el ser más poderoso en la historia de la vida en la Tierra y todos los seres vivos, incluidos los humanos, y esto podría pasar en las próximas décadas.

Si nuestros escasos cerebros fueron capaces de inventar wifi, entonces algo 100 o 1,000 o 1,000 millones de veces más inteligente que nosotros no debe tener ningún problema en controlar el posicionamiento de todos los átomos en el mundo como quiera, en cualquier momento, **todo lo que consideramos magia, todo poder que imaginamos que un Dios supremo pueda tener, podría ser una actividad tan mundana para la ASI como apretar un interruptor de luz.** La creación de la tecnología para revertir el envejecimiento humano, la cura de enfermedades, el hambre e incluso la mortalidad, reprogramar el tiempo para proteger el futuro de la vida en la Tierra, de repente todo es posible. También es posible el cese inmediato de toda la vida en la Tierra.

En lo que a nosotros respecta, si se crea un ASI, habrá una suerte de Dios omnipotente en la Tierra y la pregunta más importante para nosotros sería: **¿Será un Dios bueno?**

Lampadia