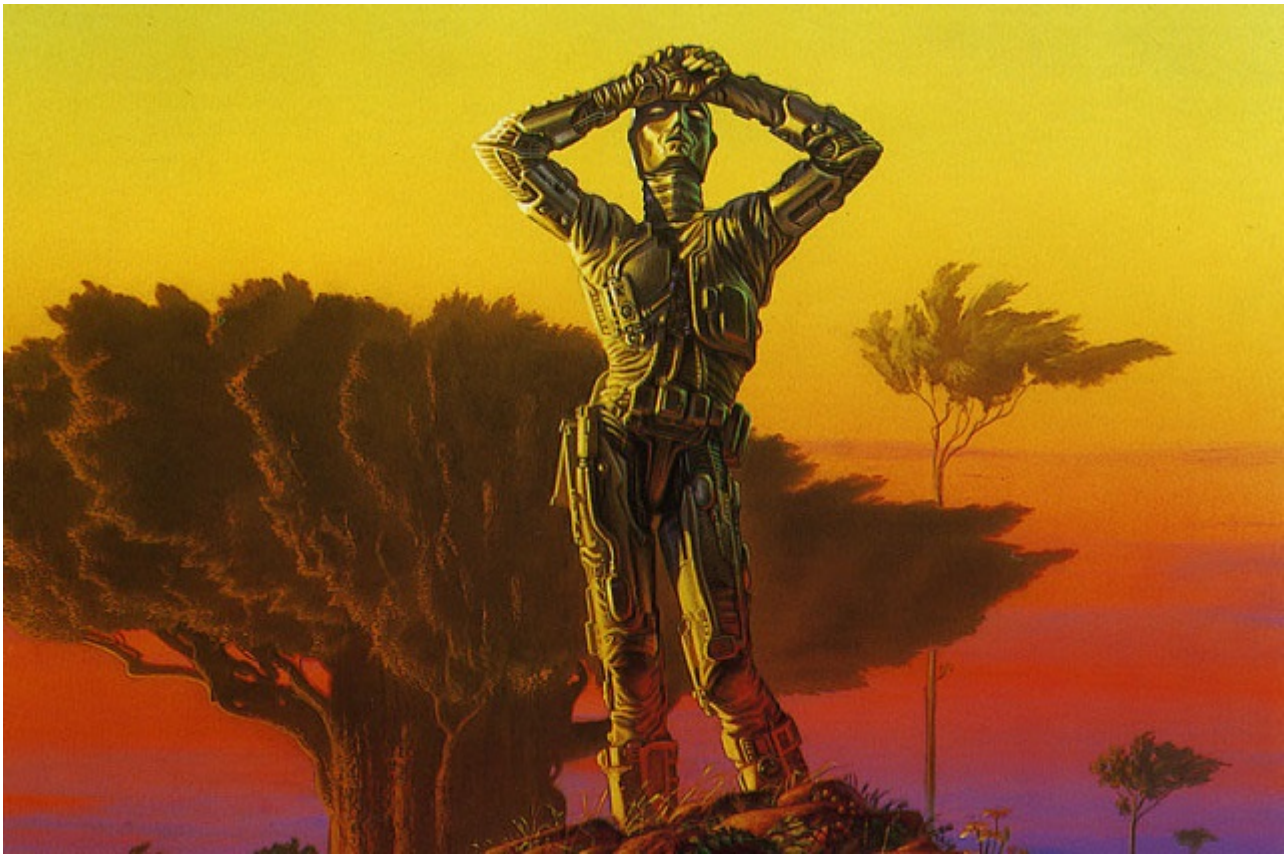
Todo lo que conoces sobre la inteligencia artificial es incorrecto



Fue recibida como la noticia más importante de la inteligencia artificial desde que Deep Blue derrotara a Gari Kaspárov al ajedrez hace casi 20 años. Google AlphaGo ha ganado tres de los cuatro primeros partidos contra el maestro Lee Sedol en un torneo de Go, lo que demuestra la espectacular mejora de la IA.

Nunca antes nos había parecido tan cercano ese fatídico día en el que las máquinas se volverán más inteligentes que los humanos, y sin embargo no llegamos a comprender las implicaciones de este acontecimiento que marcará una época. De hecho tenemos arraigada una serie de errores de concepto serios (e incluso peligrosos) sobre la inteligencia artificial.

Elon Musk, el cofundador de SpaceX, nos advirtió hace unos meses de que la IA podría acabar tomando el mundo –lo que provocó un aluvión de comentarios tanto de condena como de apoyo. Para tratarse de un evento futuro tan monumental, existe una sorprendente cantidad de desacuerdo sobre si sucederá o no, o qué forma adquirirá. Esto es especialmente preocupante si tenemos en cuenta los enormes beneficios que obtenemos de la IA y los posibles riesgos. A diferencia de cualquier otra invención humana, la IA tiene el potencial de cambiar la forma de la humanidad, pero también de destruirnos.

Es difícil saber qué creer. Sin embargo, empieza a surgir una imagen más clara gracias al trabajo pionero de los científicos de la computación, los neurocientíficos y los teóricos de la IA. Estos son los conceptos erróneos y los mitos más comunes sobre la IA.

Mito: "Nunca crearemos una IA con inteligencia similar a la humana"

Realidad: Ya tenemos máquinas que igualan o superan la capacidad humana en juegos como el ajedrez o el Go, en la compraventa del mercado de valores y en las conversaciones. Los ordenadores y los algoritmos que los impulsan sólo pueden mejorar, y será cuestión de tiempo que estas máquinas destaquen en cualquier actividad humana.

Gary Marcus, psicólogo de investigación en la Universidad de Nueva York, dijo que "prácticamente todos" los que trabajan en IA creen que las máquinas nos superarán algún día. "La única diferencia entre los entusiastas y los escépticos es el marco de tiempo". Futuristas como Ray Kurzweil dicen que podría ocurrir dentro de un par de décadas, mientras que otros creen que podría llevar siglos.

Los escépticos de la IA no resultan convincentes cuando dicen que es un problema tecnológico sin solución y que hay algo intrínsecamente único en los cerebros biológicos. Nuestros

cerebros son máquinas biológicas, pero son máquinas al fin y al cabo; existen en el mundo real y se adhieren a las leyes básicas de la física. No hay nada que sea imposible de conocer sobre ellos.

Mito: “La inteligencia artificial tendrá conciencia”

Realidad: Una suposición común en torno a la inteligencia artificial es que va a adquirir conciencia, es decir, que acabarán pensando como lo hacemos los humanos. Es más, críticos como Paul Allen, cofundador de Microsoft, creen que aún nos queda por lograr una Inteligencia Artificial Fuerte (o AGI) que sea capaz de realizar cualquier tarea intelectual que un ser humano puede hacer porque carecemos de una teoría científica de la conciencia. Pero, como señala Murray Shanahan, ingeniero de robótica cognitiva del Imperial College de Londres, debemos evitar confundir estos dos conceptos.

“La conciencia es sin duda un tema fascinante e importante, pero no creo que sea necesaria para una inteligencia artificial de nivel humano” explica Shanahan a Gizmodo. “Para ser más precisos, utilizamos la palabra conciencia para

referirnos a una serie de atributos psicológicos y cognitivos que vienen incluidos en los seres humanos”.

Es posible imaginar una máquina muy inteligente que carezca de algunos de estos atributos. Con el tiempo, podemos construir una IA que sea extremadamente inteligente, pero incapaz de experimentar el mundo de una manera consciente de sí misma, subjetiva. Murray Shanahan mencionó que podría ser posible acoplar tanto la inteligencia como la conciencia en una máquina, pero que no debemos perder de vista el hecho de que son dos conceptos distintos.

Aunque una máquina pase el test de Turing —en el que un ordenador se vuelve indistinguible de un ser humano—, eso no quiere decir que sea consciente. Para nosotros, una IA avanzada puede dar la sensación de conciencia, pero no será más consciente de sí misma que una piedra o una calculadora.

Mito: “No debemos temer a la IA”

Realidad: En enero, Mark Zuckerberg, el fundador de Facebook, dijo que no debíamos temer a la IA porque hará una cantidad increíble de cosas buenas para mejorar el mundo. Tiene razón a medias: estamos posicionados para obtener enormes beneficios de la IA —desde los coches autónomos hasta la creación de nuevos medicamentos—, pero no hay garantías de que todas las instancias de la IA serán benignas.

Un sistema altamente inteligente podría saberlo todo acerca de una tarea determinada, como por ejemplo resolver un problema financiero o hackear un sistema enemigo, pero fuera de estos ámbitos especializados ser ignorante e inconsciente. El sistema DeepMind de Google es competente en el Go, pero no tiene capacidad o raciocinio para investigar fuera de este dominio.

Muchos de estos sistemas podrían no estar desarrollados siguiendo las consideraciones de seguridad. Un buen ejemplo es

el virus Stuxnet, un gusano desarrollado por los militares de Estados Unidos e Israel para infiltrarse y atacar las plantas nucleares iraníes. De alguna manera (ya sea deliberada o accidental), este malware acabó infectando una planta de energía nuclear de Rusia.

También está Flame, un programa utilizado para el ciberespionaje en Oriente Medio. Es fácil imaginar cómo las futuras versiones de Stuxnet o Flame podrían propagarse más allá de sus objetivos e infligir incontables daños en infraestructuras sensibles.

Mito: "La súper IA será demasiado inteligente como para cometer errores"

Realidad: Richard Loosemore, matemático del Wells College, cree que los escenarios del Día del juicio final provocados por una inteligencia artificial son imposibles. Su razonamiento es que una IA lo bastante sofisticada será capaz de detectar fallos en su propio diseño y modificarse a sí misma para ser segura. Desafortunadamente, seguirá trabajando para el propósito por el que fue creada.

Peter McIntyre y Stuart Armstrong, ambos del Instituto para el Futuro de la Humanidad en la Universidad de Oxford, no están de acuerdo. Ambos creen que una IA sí que es capaz de cometer errores o simplemente puede ser demasiado necia como para saber lo que se espera de ella.

"Por definición, una superinteligencia artificial (SIA) es un agente con un intelecto muy superior al de los mejores cerebros humanos en prácticamente cualquier campo", escribe McIntyre a Gizmodo. "Sabrá exactamente lo que esperamos de ella". McIntyre y Armstrong creen que una IA sólo llevará a cabo aquellas funciones para las que está programada, pero si se vuelve lo bastante inteligente, podría ser capaz de deducir en qué difieren esas acciones del espíritu de la ley o de la intención de los seres humanos que la crearon.

McIntyre compara el futuro de los seres humanos con el de los ratones. Un ratón tiene el impulso de comer y buscar refugio, pero ese impulso a menudo entra en conflicto con los humanos, que quieren una morada libre de ratones. “Al igual que nosotros somos lo bastante inteligentes como para entender las metas de los ratones, un sistema superinteligente podría saber qué es lo que queremos y que al mismo tiempo le dé completamente igual”.

Mito: “Un simple parche solucionará el problema de controlar la IA”

Realidad: Asumiendo que podamos crear una IA más inteligente que nosotros, aún tendremos un reto muy serio conocico como “Problema del control”. Los futurólogos y teóricos de la IA no logran explicar cómo seremos capaces de contener y controlar una SIA una vez exista, o cómo lograr que sea amistosa con los seres humanos.

Recientemente, investigadores del Instituto Tecnológico de California sugirieron inocentemente que una IA podría aprender los valores y convenciones sociales de los seres humanos simplemente leyendo cuentos. Probablemente será mucho más complicado que eso.

“Se han propuesto muchos trucos sencillos que solucionarían el problema del control de la IA”, explica Armstrong. Algunos ejemplos incluyen programar la SIA de manera que tenga el impulso de satisfacer a los seres humanos, o que funcione única y exclusivamente como una herramienta. También podríamos integrar en su programación conceptos como el amor o el respeto. Para prevenir que la IA tenga una visión del mundo demasiado simplista, podría ser programada para apreciar la diversidad intelectual, cultural y social.

Sin embargo, estas soluciones son demasiado simples. Tratan de encajar toda la complejidad de los gustos y disgustos humanos en una única definición cómoda. Intentan reducir todas las

complejidades de los valores humanos en una única palabra, frase o idea. Basta con pensar en la increíble dificultad de establecer una definición coherente y práctica de términos como “respeto”.

“Eso no significa que esos parches o trucos sean completamente inútiles. Muchas de ellas sugieren líneas de investigación muy interesantes y pueden contribuir a dar con la solución definitiva”, asegura Armstrong. “Pero no podemos confiar en su eficacia sin trabajarlos mucho más y explorar todas sus implicaciones”.

Mito: “Seremos destruidos por una superinteligencia artificial”

Realidad: No hay ninguna certeza de que una IA vaya a destruirnos ni de que finalmente encontremos métodos para contenerla y controlarla. El teórico de la inteligencia artificial Eliezer Yudkowsky dice: “La IA no te odia ni te ama, pero estás hecho de átomos que puede usar con otros propósitos”.

En su libro: Superintelligence: Paths, Dangers, Strategies, el filósofo de la Universidad de Oxford Nick Bostrom escribe que una superinteligencia artificial, una vez terminada, puede suponer un peligro mayor para el ser humano que cualquier otro invento. Pensadores tan prominentes como Elon Musk, Bill Gates, o Stephen Hawking (el cual ya ha advertido que la IA puede ser el peor error de nuestra historia) también han dado la voz de alarma.

McIntyre explica que, en la mayor parte de metas que una superinteligencia artificial pueda tener, hay buenas razones para eliminar a los humanos de la ecuación.

“Una IA puede llegar a la conclusión (bastante correcta) de que no queremos que maximice el beneficio de una compañía por encima de los consumidores, el medio ambiente o los animales”,

dice McIntyre. “En ese momento, la IA tiene un ben incentivo para asegurarse de que los humanos no la interrumpen o interfieren con su objetivo, incluyendo que la apaguemos o que queramos cambiar sus metas”.

A menos que los objetivos de una SIA se correspondan exactamente con los nuestros, McIntyre explica que el sistema tendría buenas razones para no darnos la opción de detenerla, y teniendo en cuenta que su inteligencia supera ampliamente la nuestra, no tenemos mucho que hacer.

Pero no hay nada garantizado. Nadie puede saber con seguridad qué forma tomará la Inteligencia Artificial, ni como podría poner en riesgo a la humanidad. Como Musk ya ha señalado, la IA puede usarse para controlar, regular o monitorizar otras IA, o podríamos dotarlas de valores humanos o con la imposición de ser amistosa con nosotros.

Mito: “La superinteligencia artificial será amistosa”

Realidad: el filósofo Immanuel Kant creía con firmeza que la inteligencia se correlacionaba con la moralidad. En su ensayo “Singularidad: un análisis filosófico”, el neurocirujano David Chalmers tomó la famosa idea de Kant y la aplicó al auge de la superinteligencia artificial.

> *Si esto es correcto… Podemos esperar que una explosión de la inteligencia artificial conduzca también a una explosión de la moral. Podemos esperar que los sistemas (superinteligentes) serán supermorales además de superinteligentes, así que podemos asumir que serán benignos.*

La idea, sin embargo, de que la inteligencia artificial avanzada estará “iluminada” intelectualmente y será inherentemente buena no acaba de concordar. Como Armstrong indicaba, hay muchos criminales de guerra inteligentes. Una

relación entre la inteligencia y la moralidad no parece existir entre los humanos, así que cuestiona la asunción directa de que también aparecerá en otras formas de inteligencia.

“Los humanos muy inteligentes que se comportan de manera inmoral tienden a causar dolor en una escala muchísimo mayor que sus pares menos inteligentes” afirma. “La inteligencia simplemente les proporciona la habilidad de ser malos de manera más inteligente, no de transformarse en buenas personas”.

Como McIntyre explica, la habilidad de un agente de conseguir un objetivo no tiene que ver con si es un objetivo inteligente o no para empezar. “Tendríamos que ser muy afortunados para que nuestros sistemas de inteligencia artificial estuviesen dotados de manera única con la capacidad de crecer de manera moral al tiempo que crecen intelectualmente. Confiar en la suerte no es la mejor de las políticas para algo que podría determinar nuestro futuro”.

Mito: “Los riesgos de la IA y la robótica son los mismos”

Realidad: Este error es particularmente común (buenos ejemplos aquí y aquí), uno perpetuado por películas de Hollywood poco rigurosas como Terminator.

Si una superinteligencia artificial como Skynet de verdad quisiese destruir la humanidad, no usaría una serie de androides equipados con metralletas. En su lugar utilizaría medidas más eficientes como, por ejemplo, liberar una plaga biológica o instigar una plaga autorreplicante de nanobots. O podría, sin más, destruir la atmósfera. La Inteligencia Artificial es potencialmente peligrosa no por lo que implica para el futuro de la robótica sino por cómo podría invocar su presencia y devastar el mundo.

Mito: “La IA en la ciencia ficción describe con fidelidad cómo será en el futuro”

Realidad: Sí, la ciencia ficción ha sido usada por autores y futuristas para hacer predicciones durante años, pero el horizonte que dibuja la posible presencia de una superinteligencia es más oscuro. Es más, la naturaleza no humana de la IA hace que sea imposible para nosotros saber, y por tanto predecir, su forma y características.

Para que la ciencia-ficción nos entretenga como humanos, la mayoría de IAs necesitan ser similares a nosotros. “Hay todo un espectro de mentes fascinante, incluso dentro de los seres humanos. Eres diferente a tu vecino. Esta variación, con todo, no es nada comparado con todas las mentes diferentes posibles que pueden llegar a existir” amplía McIntyre.

La mayoría de ciencia-ficción existe porque necesita contar una historia con fuerza, no para ser científicamente correctas. Por tanto, el conflicto en la ciencia ficción tiende a estar entre entidades que rara vez son equiparadas. “Imagina cómo de aburrida sería una historia” dice Armstrong “donde una IA sin conciencia, felicidad u odio elimina a todos los humanos sin ningún tipo de resistencia para conseguir un objetivo que es, de por sí, poco interesante”.

Mito: “Es terrible que las IAs acaben por robarle el trabajo a humanos”

Realidad: La capacidad de la IA de automatizar muchas de las cosas que hacemos por un lado y su potencial de destruir la humanidad, por otro, son cosas muy diferentes. Según Martín Ford, autor de Rise of the Robots: Technology and the Threat of a Jobless Future, a menudo están mezclados y confundidos. Es correcto pensar en un futuro lejano y en las implicaciones que sobre él puede tener la IA, pero sólo si no nos distrae de

los problemas a los que tendremos que enfrentarnos en las siguientes décadas. El más importante de ellos es la automatización en masa.

No hay duda de que la inteligencia artificial está destinada a eliminar y reemplazar mucho de los trabajos actuales, desde el trabajo en fábricas a otros más sofisticados. Algunos expertos predicen (PDF) que la mitad de los trabajos en Estados Unidos son susceptibles de ser automatizados en el futuro.

La cuestión es que nada de esto implica que seremos incapaces de manejar la disrupción que supondrá. Está claro que eliminar mucha de nuestra carga de trabajo, tanto física como mental, es un objetivo casi utópico para nuestra especie.

“Durante las siguientes dos décadas la IA va a destruir muchos trabajos, pero eso es algo bueno” Miller explica a Gizmodo. Camiones autónomos podrían reemplazar a los transportistas, por ejemplo, lo cual abarataría los bienes provocando que fuese más barato comprar bienes “Si eres un conductor de camiones, está claro que pierdes, pero todos los demás ganan porque pueden comprar más con lo mismo. Ese dinero de los que sí ganan se podrá gastar en otros bienes y servicios que generarán nuevos trabajos para más humanos”.

Con toda probabilidad, la inteligencia artificial producirá nuevos modos de creación de riqueza, al tiempo que liberan a los humanos para realizar otras cosas. Y los avances en la IA estarán acompañados de otros en diferentes áreas, sobre todo en fabricación. En el futuro será más sencillo, y no más difícil, satisfacer nuestras necesidades básicas.

Fuente: gizmodo.