“Hay una gran presión económica para hacer obsoletos a los humanos”

En su libro *Vida 3.0* el profesor del MIT propone argumentos para un debate global que evite que la llegada de la Inteligencia Artificial acabe en desastre

DANIEL MEDIAVILLA

13 AGO 2018 - 13:42 CEST

Cuando el rey Midas le pidió a Dionisio transformar en oro todo lo que tocase cometió un fallo de programación. No pensaba que el dios sería tan literal al concederle el deseo y solo fue consciente de su error cuando vio a su hija convertida en una estatua metálica. Max Tegmark (Estocolmo, 1967) cree que la inteligencia artificial puede presentar riesgos y oportunidades similares para la humanidad.

El profesor del MIT y director del *Future of Life Institute* en Cambridge (EE UU) estima que la llegada de una Inteligencia Artificial General (IAG) que supere a la humana es cuestión de décadas. En su visión del futuro, podríamos acabar viviendo en una civilización idílica donde robots superinteligentes harían nuestro trabajo, crearían curas para todas nuestras enfermedades o diseñasen sistemas para ordeñar la energía descomunal de los agujeros negros. Sin embargo, si no somos capaces de transmitirle nuestros objetivos con precisión, también es posible que a esa nueva inteligencia dominante no le interese nuestra supervivencia o, incluso, que asuma un objetivo absurdo como transformar en clips metálicos todos los átomos del universo, los que conforman nuestros cuerpos incluidos.

Para evitar el apocalipsis, Tegmark considera que la comunidad global debe implicarse en un debate para orientar el desarrollo de la inteligencia artificial en nuestro beneficio. Esta discusión deberá afrontar problemas concretos, como la gestión de las desigualdades generadas por la automatización del trabajo, pero también un intenso esfuerzo filosófico que triunfe donde llevamos siglos fracasando y permita definir y acordar qué es bueno para toda la humanidad para después inculcárselo a las máquinas.

Estos y otros temas relacionados con la discusión que Tegmark considera más importante para el futuro de la humanidad son los que recoge en su libro *Vida 3.0: ser humano en la era de la inteligencia artificial*, un ambicioso ensayo que han recomendado gurús como Elon Musk en el que el cosmólogo sueco trata de adelantarse a lo que puede suceder durante los próximos milenios.

Pregunta. Los humanos, en particular durante los últimos dos o tres siglos, hemos tenido mucho éxito comprendiendo el mundo físico, gracias al avance de disciplinas como la física o la química, pero no parece que hayamos sido tan eficaces entendiéndonos a nosotros mismos, averiguando cómo ser felices o llegando a acuerdos sobre cómo hacer un mundo mejor para todo el mundo. ¿Cómo vamos a dirigir los objetivos de la IAG sin alcanzar antes acuerdos sobre estos asuntos?

Respuesta. Creo que nuestro futuro puede ser muy interesante si ganamos la carrera entre el poder creciente de la tecnología y la sabiduría con la que se gestiona esa tecnología. Para conseguirlo, tenemos que cambiar estrategias. Nuestra estrategia habitual consistía en aprender de nuestros errores. Inventamos el fuego, la fastidiamos

unas cuantas veces y después inventamos el extintor; inventamos el coche, la volvimos a fastidiar varias veces e inventamos el cinturón de seguridad y el *airbag*. Pero con una tecnología tan potente como las armas atómicas o la inteligencia artificial sobrehumana no vamos a poder aprender de nuestros errores. Tenemos que ser proactivos.

Es muy importante que no dejemos las discusiones sobre el futuro de la IA a un grupo de *frikis* de la tecnología como yo sino que incluyamos a psicólogos, sociólogos o economistas para que participen en la conversación. Porque si el objetivo es la felicidad humana, tenemos que estudiar qué significa ser feliz. Si no hacemos eso, las decisiones sobre el futuro de la humanidad las tomarán unos cuantos *frikis* de la tecnología, algunas compañías tecnológicas o algunos Gobiernos, que no van a ser necesariamente los mejor cualificados para tomar estas decisiones para toda la humanidad.

P. ¿La ideología o la forma de ver el mundo de las personas que desarrollen la inteligencia artificial general definirá el comportamiento de esa inteligencia?

R. Muchos de los líderes tecnológicos que están construyendo la IA son muy idealistas. Quieren que esto sea algo bueno para toda la humanidad. Pero si se mira a las motivaciones de las compañías que están desarrollando la IA, la principal es ganar dinero. Siempre harás más dinero si reemplazas humanos por máquinas que puedan hacer los mismos productos más baratos. No haces más dinero diseñando una IA que es más bondadosa. Hay una gran presión económica para hacer que los humanos sean obsoletos.

La segunda gran motivación entre los científicos es la curiosidad. Queremos ver cómo se puede hacer una inteligencia artificial por ver cómo funciona, a veces sin pensar demasiado en las consecuencias. Logramos construir armas atómicas porque había gente con curiosidad por saber cómo funcionaban los núcleos atómicos. Y después de inventarlo, muchos de aquellos científicos desearon no haberlo hecho, pero ya era demasiado tarde, porque para entonces ya había otros intereses controlando ese conocimiento.

P. En el libro parece que da por hecho que la IA facilitará la eliminación de la pobreza y el sufrimiento. Con la tecnología y las condiciones económicas actuales, ya tenemos la posibilidad de evitar una gran cantidad de sufrimiento, pero no lo hacemos porque no nos interesa lo suficiente o no le interesa a la gente con el poder necesario para conseguirlo. ¿Cómo podemos evitar que eso suceda cuando tengamos los beneficios de la inteligencia artificial?

R. En primer lugar, la tecnología misma puede ser muy útil de muchas maneras. Cada año hay mucha gente que muere en accidentes de tráfico que probablemente no morirían si fuesen en coches autónomos. Y hay más gente en América, diez veces más, que mueren en accidentes hospitalarios. Muchos de esos se podrían salvar con IA si se utilizase para diagnosticar mejor o crear mejores medicinas. Todos los problemas que no hemos sido capaces de resolver debido a nuestra limitada inteligencia es algo que podría resolver la IA. Pero eso no es suficiente. Como dice, ahora mismo tenemos muchos problemas que sabemos exactamente cómo resolver, como el hecho de que haya niños que vivan en países ricos y no estén bien alimentados. No es un problema tecnológico, es un problema de falta de voluntad política. Esto muestra lo importante de que la gente participe en esta discusión y seleccionemos las prioridades correctas.

Por ejemplo, en España, el Gobierno español ha rechazado unirse a Austria y muchos otros países en la ONU en un intento para prohibir las armas letales autónomas. España apoyó la prohibición de armas biológicas, algo que apoyaban los científicos de esa área, pero no han hecho lo mismo para apoyar a los expertos en IA. Esto es algo que la gente puede hacer: Animar a sus políticos para que afronten estos asuntos y nos aseguremos de que dirigimos la tecnología en la dirección adecuada.

P. La conversación que propone en *Vida 3.0* sobre la Inteligencia Artificial en el fondo es muy parecida a la que se debería tener sobre política en general, sobre cómo convivimos entre nosotros o como compartimos los recursos. ¿Cómo crees que el cambio en la situación tecnológica va a cambiar el debate público?

R. Creo que va a hacer las cosas más drásticas. Los cambios producidos por la ciencia se están acelerando, todo tipo de trabajos desaparecerán cada vez más rápido. Muchos se ríen de la gente que votó a Trump o a favor del Brexit, pero su rabia es muy real y los economistas te dirán que las razones por las que esta gente está enfadada, por ser más pobres de lo que eran sus padres, son reales. Y mientras no se haga nada para resolver estos problemas reales, su enfado aumentará.

La Inteligencia Artificial puede crear una cantidad enorme de nueva riqueza, no se trata de un juego de suma cero. Si nos convencemos de que va a haber suficientes impuestos para proporcionar servicios sociales y unos ingresos básicos, todo el mundo estará feliz en lugar de enfadado. Hay gente a favor de la Renta Básica Universal, pero es posible que haya mejores formas de resolver el problema. Si los gobiernos van a dar dinero a la gente solo para apoyarles, también se lo puede dar para que la gente trabaje como enfermeros o como profesoras, el tipo de trabajos que se sabe que dan un propósito a la vida de la gente, conexiones sociales...

No podemos volver a los criterios de distribución del Egipto de los faraones, en los que todo estaba en manos de un puñado de individuos, pero si una sola compañía puede desarrollar una inteligencia artificial general, es solo cuestión de tiempo que esa compañía posea casi todo. Si la gente que acumule este poder no quiere compartirlo el futuro será complicado.

P. Si no hacemos nada, ¿cuál serían las principales amenazas provocadas por el desarrollo de la IA?

R. En los próximos tres años comenzaremos una nueva carrera armamentística con armas letales autónomas. Se producirán de forma masiva por los superpoderes y en poco tiempo organizaciones como ISIS podrán tenerlas. Serán los AK-47 del futuro salvo que en este caso son máquinas perfectas para perpetrar asesinatos anónimos. En diez años, si no hacemos nada, vamos a ver más desigualdad económica. Y por último, hay mucha polémica sobre el tiempo necesario para crear una inteligencia artificial general, pero más de la mitad de los investigadores en IA creen que sucederá en décadas. En 40 años nos arriesgamos a perder completamente el control del planeta a manos de un pequeño grupo de gente que desarrolle la IA. Ese es el escenario catastrófico. Para evitarlo necesitamos que la gente se una a la conversación.

https://elpais.com/elpais/2018/08/07/ciencia/1533664021_662128.html