

EL MITO DE LA INTELIGENCIA ARTIFICIAL

Por qué las máquinas no pueden pensar
como nosotros lo hacemos

ERIK J. LARSON

Traducción de Milo J. Krmpotić

Shackleton
— b o o k s —

El mito de la Inteligencia Artificial. Por qué las máquinas no pueden pensar como nosotros lo hacemos

Título original: *The Myth of Artificial Intelligence. Why computers can't think the way we do*

© 2021 Erik J. Larson

© de esta edición, Shackleton Books, S. L., 2022

La presente edición se publica en acuerdo con Harvard University Press a través de International Editors' Co

© Traducción: Milo J. Krmpotić

Shackleton
— b o o k s —

   @Shackletonbooks
shackletonbooks.com

Realización editorial: La Letra, S. L.

Diseño de cubierta: Pau Taverna

ISBN: 978-84-1361-168-6

Depósito legal: B 14979-2022

Impreso por EGEDSA (España)

Reservados todos los derechos. Quedan rigurosamente prohibidas la reproducción total o parcial de esta obra por cualquier medio o procedimiento y su distribución mediante alquiler o préstamo públicos.

Para Brooke y Ben



CONTENIDO

Introducción	7
Primera parte: EL MUNDO SIMPLIFICADO	
1. El error de la inteligencia	15
2. Turing en Bletchley	27
3. El error de la superinteligencia	43
4. La singularidad, ayer y hoy	57
5. La comprensión del lenguaje natural	64
6. De la IA como tecnología <i>kitsch</i>	76
7. Simplificaciones y misterios	86
Segunda parte: EL PROBLEMA DE LA INFERENCIA	
8. No calcules, analiza	109
9. El puzle de Peirce (y el rompecabezas de Peirce)	116
10. Problemas de deducción e inducción	128
11. El aprendizaje automático y el <i>big data</i>	160
12. La inferencia abductiva	189
13. Inferencia y lenguaje 1	228
14. Inferencia y lenguaje 2	244

Tercera parte: EL FUTURO DEL MITO

15. Mitos y héroes	281
16. La mitología de la IA invade la neurociencia	290
17. Las teorías de la inteligencia humana basadas en el neocórtex	311
18. ¿El fin de la ciencia?	318
NOTAS	333
AGRADECIMIENTOS	349
ÍNDICE ALFABÉTICO	351

Introducción

En las páginas de este libro vas a leer acerca del mito de la inteligencia artificial. Lo de «mito» no se refiere a la imposibilidad de una IA verdadera. A ese respecto, el futuro de la IA es un misterio para la ciencia. El mito de la inteligencia artificial consiste en afirmar que su llegada es inevitable, mera cuestión de tiempo —que nos hemos adentrado ya en el sendero que conducirá a una IA de nivel humano, y más tarde a una superinteligencia—. No es así. Ese sendero existe solo en nuestra imaginación. Sin embargo, el carácter inevitable de la IA se encuentra tan arraigado en el debate popular —promovido por los expertos de los medios de comunicación, por referentes intelectuales como Elon Musk e incluso por numerosos científicos de IA (aunque desde luego no por todos ellos)— que, a menudo, cualquier pega que se le ponga se considera una forma de ludismo, o por lo menos una visión corta de miras sobre el futuro de la tecnología y un fracaso peligroso a la hora de prepararse para un mundo de máquinas inteligentes.

Tal y como os voy a mostrar, la ciencia de la IA ha revelado un misterio de grandes dimensiones en el núcleo de la inteligencia, y en la actualidad nadie tiene la menor idea de cómo resolverlo. Los partidarios de la IA cuentan con inmensos incentivos para minimizar sus limitaciones. Al fin y al cabo, la IA es un negocio enorme y tiene una presencia cada vez más predominante en la cultura. No obstante, nos guste o no, la posibilidad de un futuro de sistemas de IA se encuentra limitada por lo que sabemos en la

actualidad sobre la naturaleza de la inteligencia. Y aquí deberíamos afirmarlo con franqueza: todas las pruebas sugieren que las inteligencias humana y artificial son radicalmente diferentes. El mito de la IA insiste en que esas diferencias son solo temporales, y en que la aparición de sistemas más potentes acabará por erradicarlas. Futurólogos como Ray Kurzweil y el filósofo Nick Bostrom, prominentes proveedores del mito, hablan no solo como si la IA de nivel humano resultara inevitable, sino como si, al poco de su llegada, las máquinas superinteligentes fueran a dejarnos muy atrás.

Este libro explica dos aspectos importantes del mito de la IA, uno de tipo científico y otro cultural. La parte científica del mito asume que solo tenemos que seguir «desnudando la cebolla» del desafío de la inteligencia general, avanzando en hitos restrictivos de la inteligencia como la participación en juegos o el reconocimiento de imágenes. Se trata de un error grave: el éxito en las aplicaciones débiles no nos acerca ni un solo paso a la inteligencia general. Las inferencias que requieren los sistemas de cara a alcanzar una inteligencia general —leer el periódico, o mantener una conversación elemental, o ejercer de ayudante, como el robot Rosie de *Los Supersónicos*— no se pueden programar, aprender ni diseñar a partir de nuestro conocimiento actual de la IA. Al aplicar con éxito versiones de inteligencia más simples y débiles, que se benefician del uso de ordenadores más rápidos y de montones de datos, no estamos obteniendo un avance progresivo, sino que nos limitamos a recoger sus frutos maduros. El salto hacia un «sentido común» general es completamente diferente, y no se conoce camino alguno que lleve de lo uno a lo otro. No existe ningún algoritmo para la inteligencia general. Y tenemos buenos motivos para mostrarnos escépticos ante la idea de que dicho algoritmo vaya a surgir de nuevas tentativas con los sistemas de aprendizaje profundo o de cualquier otra aproximación popular en la actualidad. Resulta mucho más probable que vaya a requerir de un avance científico de primer orden, y ahora mismo nadie tiene la más remota idea del aspecto que tendría ese avance, y mucho menos de los detalles que conducirán a él.

La mitología sobre la IA es negativa, pues, porque oculta un misterio científico bajo la cháchara interminable del progreso continuado. El mito sostiene la creencia en un éxito inevitable, pero el respeto genuino por la ciencia debería hacer que volviéramos a la casilla de salida. Eso nos conduce al segundo tema de estas páginas: las consecuencias culturales del mito. Perseguir un mito no es la mejor manera de obtener «inversiones expertas», y ni siquiera una posición neutral. Es malo para la ciencia y es malo para nosotros. ¿Por qué? Un motivo es que resulta poco probable que alcancemos innovaciones si decidimos ignorar un misterio tan básico en vez de afrontarlo. La versión saludable de la cultura de la innovación pone el énfasis en la exploración de lo que se desconoce, no en dar bombo a la ampliación de unos métodos ya existentes —sobre todo cuando esos métodos se han revelado inadecuados para llevarnos mucho más allá—. La mitología acerca del éxito inevitable de la IA tiende a extinguir la cultura misma de la invención, tan necesaria para obtener un avance real —con la IA de nivel humano o sin ella—. El mito también fomenta la resignación ante el progresivo avance hacia una tierra de máquinas, donde la invención genuina se deja de lado en favor de charlas futuristas que defienden los métodos actuales, a menudo desde intereses particulares.

¿Quién debería leer este libro? Sin duda, cualquier persona que se emocione con la idea de la IA pero que se esté preguntando por qué siempre aparece a diez o veinte años vista. Hay un motivo científico para ello, que explicaré. También deberías leer este libro si piensas que el progreso de la IA hacia la superinteligencia es inevitable y te preocupa lo que habrá que hacer cuando llegue. Aunque no puedo demostrar que una *autoridad suprema* de la IA no vaya a aparecer algún día, sí puedo ofrecerte razones para que descartes con rigurosidad la perspectiva de ese escenario. Más en general, debes leer este libro si sientes curiosidad pero a la vez te encuentras confundido por el bombo generalizado que rodea a la IA en nuestra sociedad. Te explicaré los orígenes del mito de la IA, lo que sabe-

mos y lo que ignoramos acerca de la perspectiva de alcanzar una IA de nivel humano, y el motivo por el que deberíamos apreciar mejor la única inteligencia verdadera que conocemos: la nuestra.

EN ESTE LIBRO

En la primera parte, «El mundo simplificado», explico que la cultura de la IA nos ha llevado a simplificar nuestras ideas sobre la gente a la vez que expandía nuestro conocimiento acerca de la tecnología. Esto comenzó con el fundador de la IA, Alan Turing, e incluye una serie de simplificaciones comprensibles pero desafortunadas que yo denomino «errores de inteligencia». Esos errores iniciales fueron magnificados hasta acabar conformando una ideología por parte de un amigo de Turing, el estadístico I. J. Good, quien introdujo la idea de «ultrainteligencia» como resultado predecible tras la consecución de una IA de nivel humano. Entre Turing y Good vemos cobrar forma al mito moderno de la IA. Su desarrollo nos ha conducido a una época de lo que yo llamo «tecnología *kitsch*», imitaciones baratas de ideas más profundas que anulan el compromiso inteligente y debilitan nuestra cultura. Lo *kitsch* nos indica lo que hemos de pensar y lo que hemos de sentir. Los proveedores del *kitsch* sacan rédito de él, mientras que los consumidores de ese *kitsch* experimentan una pérdida; acaban —acabamos— metidos en un mundo de frivolidad.

En la segunda parte, «El problema de la inferencia», argumento que no tenemos la menor idea sobre cómo programar o diseñar el único tipo de inferencia —de pensamiento, en otras palabras— que funcionará con una IA de nivel humano (o cualquier otra cosa que se le acerque). El problema de la inferencia apunta al corazón del debate sobre la IA porque trata directamente con la inteligencia, la de la gente o la de las máquinas. Nuestro conocimiento acerca de los distintos tipos de inferencia se remonta a Aristóteles y a otros griegos de la Antigüedad, y se ha desarrolla-

do en los ámbitos de la lógica y de las matemáticas. La inferencia ya se describe usando sistemas formales y simbólicos como los programas informáticos, así que explorándola se puede obtener una visión muy clara del proyecto con el que diseñar la inteligencia. Hay tres tipos de inferencia. La IA clásica exploró uno (las deducciones), la IA moderna explora otro (las inducciones). Y el tercer tipo (las abducciones) conduce a la inteligencia general y, sorpresa: nadie está trabajando en él —nadie en absoluto—.¹ Por último, puesto que todos los tipos de inferencia son distintos —con ello quiero decir que ninguno de esos tipos puede rebajarse hasta convertirse en otro—, sabemos que un fracaso a la hora de construir sistemas de IA que usen el tipo de inferencia en el que se afianza la inteligencia general conducirá al fracaso de los avances hacia la inteligencia artificial general, o IAG.

En la tercera parte, «El futuro del mito», argumento que, cuando se lo toma uno en serio, el mito tiene consecuencias muy negativas, ya que subvierte la ciencia. En especial, erosiona la cultura de la invención y la inteligencia humanas, que resultan necesarias en aquellos descubrimientos imprescindibles para comprender nuestro propio futuro. La ciencia de datos (la aplicación de la IA a los macrodatos) es, en el mejor de los casos, una prótesis del ingenio humano; en caso de usarla de manera correcta, nos ayudará a lidiar con el «diluvio de datos» contemporáneo. Cuando se la usa para reemplazar la inteligencia individual, tiende a estropear la inversión sin ofrecer ningún resultado. Explico, en especial, que el mito ha afectado negativamente la investigación en neurociencia, entre otros avances científicos recientes. Estamos pagando un precio demasiado elevado por este mito. Como no poseemos ninguna buena razón científica para creer que el mito pueda hacerse realidad, puesto que contamos con todos los motivos para rechazarlo a fin de alcanzar la prosperidad en el futuro, tenemos que repensar de manera radical la conversación sobre la IA.