



ALIBABA

IBM

GOOGLE

MICROSOFT

# LOS NUEVE GIGANTES

AMAZON

FACEBOOK

Cómo las grandes  
tecnológicas amenazan  
el futuro de la humanidad

AMY WEBB

TENCENT

BAIDU

APPLE

# Los nueve gigantes

Cómo las grandes tecnológicas amenazan  
el futuro de la humanidad

Amy Webb

Traducción de María Mercedes Correa

Título original: *The Big Nine: How the Tech Titans and Their Thinking Machines Could Warp Humanity*

© Amy Webb, 2019

«This edition published by arrangement with Public Affairs, an imprint of Perseus Books, LLC a subsidiary of Hachette Book Group Inc., New York, USA. All rights reserved»

Queda rigurosamente prohibida sin autorización por escrito del editor cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra, que será sometida a las sanciones establecidas por la ley. Pueden dirigirse a Cedro (Centro Español de Derechos Reprográficos, [www.cedro.org](http://www.cedro.org)) si necesitan fotocopiar o escanear algún fragmento de esta obra ([www.conlicencia.com](http://www.conlicencia.com);

91 702 19 70 / 93 272 04 47).  
Todos los derechos reservados.

Primera edición: junio de 2021

© de la traducción del inglés: María Mercedes Correa, 2020

© de esta edición: Edicions 62, S.A., 2021  
Ediciones Península,  
Diagonal 662-664  
08034 Barcelona  
[edicionespeninsula@planeta.es](mailto:edicionespeninsula@planeta.es)  
[www.edicionespeninsula.com](http://www.edicionespeninsula.com)

DAVID PABLO - fotocomposición  
DEPÓSITO LEGAL: B. 7.586-2021  
ISBN: 978-84-9942-995-3

## ÍNDICE

Introducción. <i>Antes de que sea demasiado tarde</i>	13
---	----

### PRIMERA PARTE ESPÍRITUS EN LA MÁQUINA

1. La mente y las máquinas: una breve historia de la inteligencia artificial	29
2. Las islas de las tribus de la inteligencia artificial	81
3. Mil cortes hechos con papel: las consecuencias no buscadas de la inteligencia artificial	143

### SEGUNDA PARTE NUESTROS POSIBLES FUTUROS

4. De aquí hasta la superinteligencia artificial: las señales de alerta	191
5. El progreso en la tercera era de la informática: el escenario optimista	219
6. Aprender a vivir con millones de cortes hechos con papel: el escenario pragmático	251
7. La dinastía Réngōng Zhìnéng: el escenario catastrófico	289

Tercera parte  
Resolver los problemas

8. Guijarros y rocas: cómo mejorar el futuro de la inteligencia artificial	321
Agradecimientos	357
Bibliografía	361
Primeros investigadores en inteligencia artificial	369
Índice onomástico	373

## LA MENTE Y LAS MÁQUINAS: UNA BREVE HISTORIA DE LA INTELIGENCIA ARTIFICIAL

Los orígenes de la inteligencia artificial (IA) se remontan a cientos de años, mucho antes de que los nueve gigantes hubieran creado agentes de IA con nombres de personas, como Siri, Alexa y su colega china Tiān Mǎo. En este largo lapso, no se ha producido una definición singular de IA, como sí la hay para otras tecnologías. No es fácil definir de una manera concreta la IA, pues representa muchas cosas, y más si se tiene en cuenta su constante crecimiento. Aquello que en la década de 1950 se podía considerar como IA (una calculadora capaz de hacer divisiones muy complejas), hoy no parece una tecnología avanzada. A esta realidad se le ha dado el nombre de «extraña paradoja»: cuando las nuevas técnicas inventadas pasan a ser de uso corriente, se vuelven invisibles para nosotros. Ya no consideramos que esa tecnología pueda ser IA.

En su forma más básica, la IA es un sistema que toma decisiones autónomas. Las tareas que lleva a cabo replican o imitan los actos de la inteligencia humana, tales como reconocer sonidos y objetos, resolver problemas, comprender el lenguaje y usar estrategias para alcanzar metas. Algunos sistemas de IA son enormes y llevan a cabo millones de operaciones en un tiempo muy breve, mientras

que otros son limitados y hacen una tarea única, como identificar el lenguaje soez en los correos electrónicos.

Siempre estamos volviendo a las mismas preguntas: ¿pueden pensar las máquinas?, ¿qué significa que una máquina «piense»? ¿qué significa que los humanos pensemos?, ¿cómo podríamos saber (de manera definitiva y sin lugar a dudas) que estamos pensando pensamientos originales? Estas preguntas nos acompañan desde hace siglos, y son fundamentales en la historia y el futuro de la IA.

El problema que surge al investigar cómo piensan las máquinas y los humanos tiene que ver con el hecho de que la palabra «pensar» está estrechamente ligada a la palabra «mente». En el diccionario *Merriam-Webster* se define «pensar» como «formar o tener en la mente», mientras que el diccionario *Oxford* explica que «pensar» significa «usar la mente de manera activa para formar ideas conectadas». Si buscamos tanto en el *Merriam-Webster* como en el *Oxford* vemos que «mente» se define dentro del contexto de la conciencia. Ahora bien, ¿qué es la conciencia? Según ambos diccionarios, es la capacidad para reconocer el entorno o la propia existencia. La idea de pensamiento varía según el enfoque de cada grupo de especialistas: psicólogos, neurocientíficos, filósofos, teólogos o expertos en ética e informática.

Cuando usted usa a Alexa para reservar una mesa en su restaurante favorito, tanto usted como ella son conscientes del intercambio sobre la comida, aunque Alexa jamás haya sentido la textura de una manzana crujiente en los dientes, ni la efervescencia de las burbujas de un refresco en la lengua, ni la viscosidad de la mantequilla de cacahuete pegada en el paladar. Si se le pide a Alexa que describa las cualidades de esos alimentos, ofrecerá detalles que se asemejan a las experiencias que usted ha tenido. Alexa no tiene boca, entonces ¿cómo podría percibir la comida como la percibe usted?

Usted es una persona biológicamente única, cuyas glándulas salivares y papilas gustativas no están distribuidas exactamente en

el mismo orden que las mías. Sin embargo, tanto usted como yo hemos aprendido qué es una manzana y cuáles son sus características de sabor, textura y olor. A lo largo de nuestra vida, hemos aprendido a reconocer qué es una manzana mediante el aprendizaje por refuerzo: alguien nos enseñó qué aspecto tiene una manzana, para qué la usamos y qué la diferencia de otras frutas. Luego, con el tiempo y sin la participación de la conciencia, nuestros sistemas biológicos autónomos de reconocimiento de patrones adquirieron una excelente habilidad para determinar si algo era una manzana, aunque solo tuviéramos algunos de los datos necesarios. Si usted ve una representación en blanco y negro y en dos dimensiones de una manzana, usted sabe lo que es, aunque falten el sabor, el olor, la textura y los demás datos que le indiquen a su cerebro «esto es una manzana». La manera en que usted y Alexa aprendieron qué es una manzana se parece más de lo que usted se imaginaba.

Alexa es competente, pero ¿es inteligente? ¿Debe su percepción mecánica reunir todas las cualidades de la percepción humana para que aceptemos que su manera de «pensar» es un reflejo de la nuestra? El psicólogo educativo Benjamin Bloom pasó la mayor parte de su vida académica investigando y clasificando los estados del pensamiento. En 1956, publicó una obra que pasó a ser conocida como la taxonomía de Bloom, en la que hace una clasificación de los objetivos de aprendizaje y de los distintos logros observados en la educación. La base consiste en recordar hechos y conceptos básicos, a lo que sigue, en su clasificación, comprender ideas; aplicar el conocimiento en situaciones nuevas; analizar la información mediante la experimentación y hacer conexiones; evaluar, defender y juzgar la información; y, por último, producir un trabajo original. Los bebés se concentran inicialmente en recordar y comprender. Por ejemplo, primero debemos saber que el biberón tiene leche, antes de comprender que ese biberón tiene una parte delantera y una posterior, aunque no podamos verlas.



Esta jerarquía también está presente en la manera en que aprenden los ordenadores. En 2017, un sistema de IA llamado Amper compuso y produjo música original para un álbum llamado *I AM AI* [Yo soy IA]. Las estructuras de los acordes, la instrumentación y la percusión fueron desarrolladas por Amper, que usó parámetros iniciales tales como el género, el modo y la longitud para generar canciones completas en tan solo unos minutos. Taryn Southern, un artista humano, colaboró con Amper en la creación del álbum, cuyo resultado incluye una balada muy expresiva llamada *Break Free* [Libérate], que tuvo más de 1,6 millones de visitas en YouTube y fue un éxito en la radio tradicional. Antes de que Amper pudiera crear esa canción, primero tuvo que aprender los elementos cualitativos de la balada, además de datos cuantitativos tales como el valor de las notas y los ritmos, así como la manera de reconocer miles de patrones en la música (por ejemplo, las progresiones de acordes, las secuencias armónicas y los acentos rítmicos).

La creatividad, esa que ilustra el ejemplo de Amper, se encuentra en el pináculo de la taxonomía de Bloom, pero cabe preguntarse si se trata tan solo un proceso mecánico aprendido. ¿Es un ejemplo de creatividad humana? ¿Es, acaso, un tipo de creatividad totalmente diferente? ¿Pensó Amper en la música de la misma manera en que lo haría un compositor humano? Podría argüirse que el «cerebro» de Amper (una red neuronal que usa algoritmos y datos dentro de un contenedor) tal vez no sea tan diferente del cerebro de Beethoven, hecho de neuronas orgánicas, que usan datos y reconocen patrones dentro de un contenedor al que llamamos «cabeza». ¿Fue el proceso creativo de Amper verdaderamente distinto del de Beethoven cuando compuso la famosa *Quinta sinfonía*, esa que comienza con la secuencia ta TAA, ta TAA, antes de pasar de una clave mayor a una menor? Beethoven no se inventó la totalidad de la sinfonía: esta no es completamente original. Las primeras cuatro notas van seguidas de una secuencia armónica, partes de escalas, arpeggios y

otros ingredientes comunes que conforman cualquier composición. Si se escucha con atención el *scherzo*, antes del final, se identificarán patrones obvios tomados de la *Sinfonía n.º 40* de Mozart, compuesta veinte años antes, en 1788. Mozart fue influenciado por su rival Antonio Salieri y por su amigo Franz Joseph Hayden, quienes, a su vez, fueron influenciados por el trabajo de compositores que los precedieron, tales como Johann Sebastian Bach, Antonio Vivaldi y Henry Purcell, compositores de los siglos XVI y XVII. Es posible detectar temas de compositores aún más antiguos, de años que van entre 1400 y 1600, como Jacques Arcadelt, Jean Mouton y Johannes Ockeghem, en su música. A su vez, ellos recibieron la influencia de los compositores medievales, y así podemos seguir buscando el patrón de influencia hasta las primeras composiciones escritas, como la del «epitafio de Sículo», grabada en una columna de mármol para una tumba turca del siglo I. Podríamos, incluso, retroceder más en el tiempo, hasta la elaboración de las primeras flautas hechas con huesos y marfil, hace 43.000 años. Y antes de eso, los investigadores creen que nuestros primeros ancestros cantaban antes de hablar.<sup>1</sup>

Nuestro cableado humano es el resultado de millones de años de evolución. De manera similar, el cableado de la IA moderna se basa en un largo recorrido evolutivo que se remonta a los antiguos matemáticos, filósofos y científicos. Aunque parezca que las máquinas y los humanos hemos recorrido caminos diferentes, nuestra evolución siempre ha estado interrelacionada. El *Homo sapiens* aprendió del entorno, transmitió sus características a las siguientes generaciones, se diversificó y se multiplicó gracias a la invención de tecnologías avanzadas, como la agricultura, los utensilios para la caza y la penicilina. Se necesitaron 11.000 años para que los

---

1 «The Seikilos Epitaph: The Oldest Song in the World», *Wired*, 29 de octubre de 2009, <<https://www.wired.com/2009/10/the-seikilos-epitaph>>.

seis millones de habitantes de la Tierra durante el Neolítico se propagaran y se convirtieran en una población de 7.000 millones de seres humanos hoy en día.<sup>2</sup>

El ecosistema donde viven los sistemas de IA (los insumos para el aprendizaje, los datos, los algoritmos, los procesadores, las máquinas y las redes neurales) se está mejorando e iterando a velocidades exponenciales. A los sistemas de IA les llevará tan solo algunas décadas propagarse e integrarse en todas las facetas de la vida cotidiana.

Preguntarse si Alexa percibe una manzana de la misma manera como lo hacen los humanos, o preguntarse si la música original de Amper es verdaderamente «original» equivale, en realidad, a preguntarse sobre la manera en que los humanos pensamos sobre el pensamiento. La IA de nuestros días es una amalgama de miles de años de filósofos, matemáticos, científicos, expertos en robótica, artistas y teólogos. Su búsqueda (y la nuestra, en el presente capítulo) apunta a la comprensión del lazo que une el pensamiento con los contenedores del pensamiento. ¿Cuál es la conexión entre la mente humana y las máquinas producidas por los nueve gigantes en China y en Estados Unidos?

#### ¿ESTÁ LA MENTE DENTRO DE UNA MÁQUINA?

El fundamento de la IA se remonta a la antigua Grecia y a los orígenes de la filosofía, la lógica y la matemática. En muchos de los escritos de Platón, Sócrates exhorta: «Conócete a ti mismo», con lo cual quería expresar que, para mejorar y tomar las decisiones correctas, es necesario conocer, en primer lugar, el carácter de uno mismo. Por otra parte, Aristóteles creó la lógica de los silogismos y nuestro primer sistema formal de razonamiento deductivo. Más o menos hacia la misma época, el matemático griego Euclides concibió una forma

---

<sup>2</sup> «Population Clock: World», *Census.gov*, 2018, <<https://www.census.gov/popclock/world>>.

de encontrar el máximo común divisor de dos números y, de esta manera, creó el primer algoritmo.

El trabajo de estos griegos dio pie a la concepción de dos importantes ideas nuevas: que ciertos sistemas físicos pueden operar como un conjunto de reglas lógicas y que el pensamiento humano mismo podría ser un sistema simbólico. Este fue el punto de partida para cientos de años de investigaciones adelantadas por filósofos, teólogos y científicos. ¿Era el cuerpo una máquina compleja? ¿Era, acaso, un todo unificado hecho de cientos de sistemas que trabajaban juntos, igual que un reloj de pie? ¿Qué decir de la mente? ¿Era, también, una máquina compleja, o era algo totalmente diferente? No había manera de probar la existencia o no de un algoritmo divino ni la conexión entre la mente y lo físico.

En 1560, un relojero español llamado Juanelo Turriano creó un monje mecánico diminuto, como ofrenda para la Iglesia, a nombre del rey Felipe II de España, cuyo hijo se había recuperado milagrosamente de una herida en la cabeza.<sup>3</sup> Esta creación tenía unos poderes sorprendentes: caminaba por las mesas, levantaba un crucifijo y un rosario, se daba golpes en el pecho a modo de contrición y movía los labios como si estuviera orando. Fue el primer autómatas, es decir, una representación mecánica de un ser viviente. Aunque la palabra robot todavía no existía, la invención del monje era maravillosa, y debía sorprender y confundir a quienes la observaban. Es probable que a nadie en aquella época se le hubiera ocurrido que un diminuto autómatas pudiera algún día, en el futuro lejano, no solamente imitar unos movimientos básicos, sino que pudiera reemplazar a los humanos en las plantas de producción, en laboratorios de investigación y en conversaciones culinarias.

---

3 Elizabeth King, «Clockwork Prayer: A Sixteenth-Century Mechanical Monk», *Blackbird* 1, n.º 1 (primavera de 2002), <[https://blackbird.vcu.edu/v1n1/nonfiction/king\\_e/prayer\\_introduction.htm](https://blackbird.vcu.edu/v1n1/nonfiction/king_e/prayer_introduction.htm)>.

El pequeño monje mecánico inspiró a la primera generación de expertos en robótica, cuyo objetivo era crear máquinas cada vez más complejas que fueran un reflejo de los humanos: los autómatas podían escribir, bailar y pintar. Esto condujo a un grupo de filósofos a preguntarse qué significa ser humano. Si era posible construir autómatas que pudieran imitar el comportamiento humano, ¿somos los humanos unos autómatas producto de una construcción divina? ¿O somos, acaso, sistemas complejos capaces de razonar y producir pensamiento original?

En *De Corpore* —una parte de la gran trilogía de Thomas Hobbes sobre las ciencias naturales, la psicología y la política—, este filósofo político inglés describió el razonamiento humano como «computación». En 1655 escribió: «Cuando hablo de razonar me refiero a computar. Y computar es hacer la suma de muchas cosas puestas juntas al mismo tiempo, o saber qué queda cuando una cosa se le sustrae a otra. Razonar, por lo tanto, es lo mismo que sumar o restar».<sup>4</sup> Ahora bien, ¿cómo podemos saber que tuvimos libre albedrío durante el proceso?

Mientras Hobbes escribía la primera parte de su trilogía, el filósofo francés René Descartes publicaba *Meditaciones de filosofía primera*, libro en el cual se preguntaba si podemos saber con certeza que lo percibido es real. ¿Cómo podemos verificar nuestra propia conciencia? ¿Qué prueba necesitaríamos para concluir que nuestros pensamientos son nuestros y que el mundo que nos rodea es real? Descartes era un racionalista que creía que podíamos aprehender los hechos por medio de la deducción. Recordemos su famoso experimento de pensamiento: Descartes les pedía a los lectores imaginar que un demonio había creado, a propósito, una ilusión del mundo. Si la experiencia física, sensorial del lector de estar nadando en un lago no era más que la construcción de un demonio,

---

<sup>4</sup> Thomas Hobbes, *De Corpore Politico, or The Elements of Law Moral and Politick*.

entonces el lector no podía saber realmente que estaba nadando. Sin embargo, en opinión de Descartes, si el lector tenía una conciencia personal de su propia existencia, cumplía el criterio que se aplica para el conocimiento. Descartes escribió: «La proposición “yo soy, yo existo” es necesariamente verdadera cada vez que la pronuncio o la concibo en mi mente». <sup>5</sup> En otras palabras, el hecho de nuestra existencia está más allá de toda duda, aun si hay de por medio un demonio que nos engaña. O sea, «Pienso, luego existo».

Más adelante, en su *Tratado del hombre*, Descartes arguye que los humanos podrían crear un autómatas (en este caso, un pequeño animal) que no sería posible distinguir del ser real. Sin embargo, continúa diciendo Descartes, si algún día creáramos un humano mecanizado, este nunca podría pasar por un humano real, porque carecería de mente y, por lo tanto, de alma. A diferencia de los humanos, una máquina nunca pasaría los criterios del conocimiento: nunca podría tener conciencia de sí misma, como la tenemos nosotros. Para Descartes, la conciencia era algo que ocurría internamente: el alma era el espíritu en la máquina que es nuestro cuerpo. <sup>6</sup>

Algunas décadas más adelante, el matemático y filósofo alemán Gottfried Wilhelm von Leibniz examinó la idea de que el alma humana estaba programada, y argüía que la mente misma era un contenedor. Dios creó el alma y el cuerpo para que se armonizaran naturalmente. Puede que el cuerpo sea una máquina compleja, pero es una máquina que viene con un conjunto de instrucciones divinas. Nuestras manos se mueven cuando decidimos moverlas, pero nosotros no creamos ni inventamos los mecanismos que permiten el movimiento. Si somos conscientes del dolor o del placer, esas sensaciones son el resultado de un sistema preprogramado, una línea continua de comunicación entre la mente y el cuerpo.

---

<sup>5</sup> René Descartes, *Meditations on First Philosophy*, Second Meditation §25, 1641, University of Connecticut, <<http://selfpace.uconn.edu/class/percep/DescartesMeditations.pdf>>.

<sup>6</sup> René Descartes, *Treatise of Man*, trad. T. S. Hall (Cambridge, MA: Harvard University Press, 1972).

Leibniz concibió su propio experimento para ilustrar el concepto de que el pensamiento y la percepción están íntimamente ligados al ser humano. Imagine el lector que entra en un molino. El edificio es un contendedor donde hay máquinas, materias primas y trabajadores. Es un complejo sistema de partes que funcionan armoniosamente para alcanzar un objetivo en particular, pero nunca podría tener una mente. «Todo lo que podríamos encontrar allí son ruedas dentadas y palancas moviéndose unas a otras, sin nada que pueda llamarse percepción», escribió Leibniz. «Así, la percepción puede encontrarse en sustancias simples, pero no en compuestos como las máquinas.» Su argumento apuntaba a que, sin importar cuán avanzado fuera el molino, la maquinaria o el autómeta, los humanos no podrían construir una máquina capaz de pensar o de percibir.<sup>7</sup>

Con todo, Leibniz sentía fascinación por la idea de la replicación de ciertas facetas del pensamiento. Algunas décadas antes, un escritor inglés poco conocido, llamado Richard Braithwaite, quien escribió sobre la conducta social, se refirió a los «computadores» humanos como personas altamente entrenadas, que podían hacer cálculos de manera muy rápida y precisa.<sup>8</sup> Entre tanto, el matemático francés Blaise Pascal, quien sentó las bases para lo que hoy conocemos como probabilidad, reflexionaba sobre la automatización de ciertas tareas computacionales. Pascal observaba cómo su padre calculaba los impuestos a mano, con gran monotonía, y quería facilitarle la tarea. Pascal comenzó a trabajar en una calculadora automática, con ruedas mecánicas y discos móviles.<sup>9</sup> La calculadora funcionaba, e inspiró a Leibniz para refinar

---

7 Gottfried Wilhelm Leibniz, *The Monadology*, trad. Robert Latta, (1898), <<https://www.plato-philosophy.org/wp-content/uploads/2016/07/The-Monadology-1714-by-Gottfried-Wilhelm-LEIBNIZ-1646-1716.pdf>>.

8 Se cree que el primer uso conocido de la palabra «computador» aparece en un libro llamado *The Yong Mans Gleanings*, escrito por Richard Braithwaite en 1613. En esa época, los computadores eran personas que hacían cálculos.

9 «Blaise Pascal», Biography.com, <<https://www.biography.com/people/blaise-pascal-9434176>>.

su pensamiento: las máquinas nunca tendrían alma, pero algún día sería posible crear una máquina capaz de tener pensamiento lógico de nivel humano. En 1673, Leibniz describió su «contador de pasos», un nuevo tipo de máquina calculadora que tomaba decisiones usando un sistema binario.<sup>10</sup> La máquina se parecía en cierto sentido a una mesa de billar, con bolas, agujeros, palos y canales. Esta máquina abría los agujeros usando una serie de ceros (cerrado) y de unos (abierto).

El contador de pasos teórico de Leibniz sentó las bases para nuevas teorías, que comprendían la siguiente idea: si el pensamiento lógico podía reducirse a símbolos y, en consecuencia, podía analizarse como un sistema computacional, y si los problemas geométricos podían resolverse usando símbolos y números, entonces todo podría reducirse a bits, incluyendo el comportamiento humano. Era un distanciamiento significativo de los anteriores filósofos: las máquinas del futuro podrían replicar los procesos del pensamiento humano sin tener que ver con la divina providencia. El pensamiento no requería necesariamente de percepción, sentidos o alma. Leibniz imaginó un ordenador capaz de resolver problemas generales, incluso de carácter no matemático. Lanzó la hipótesis de que el lenguaje podía reducirse a los conceptos atómicos de las matemáticas y las ciencias, como parte de un traductor de lenguaje universal.<sup>11</sup>

---

10 Leibniz escribió en *De progressionē dyadica*: «Este cálculo binario podría ser implementado por una máquina [...] provista de agujeros, que puedan abrirse y cerrarse. Deben estar abiertos en los lugares que corresponden a 1 y permanecer cerrados en los que corresponden a 0. Por los agujeros abiertos entran unas canicas, que ruedan por unos canales; por los cerrados, nada. La disposición de los agujeros debe cambiarse de columna en columna, según se requiera».

11 Leibniz escribió: «Pensé nuevamente en mi plan para un nuevo lenguaje o sistema de razonamiento mediante la escritura, que podría servir como herramienta de comunicación entre todas las diferentes naciones [...]. Si tuviéramos esa herramienta universal, podríamos discutir problemas de metafísica o cuestiones éticas de la misma manera como lo hacemos con los problemas o las cuestiones de matemáticas o geometría. Ese era mi propósito: cualquier malentendido no sería más que un error de cálculo [...], fácilmente corregible mediante las leyes gramaticales de ese nuevo lenguaje. Así, en el caso de una discusión controvertida, dos filósofos podrían sentarse en una mesa y calculando, como dos matemáticos, simplemente dirían: "revisemos"».



¿LA MENTE Y LAS MÁQUINAS SIMPLEMENTE SIGUEN UN ALGORITMO?

Si Leibniz estaba en lo cierto (los humanos eran máquinas con alma, que algún día inventarían máquinas sin alma, capacitadas para tener un inimaginable pensamiento sofisticado), entonces habría sobre la Tierra una clase binaria de máquina: ellas y nosotros. Pero este era tan solo el comienzo del debate.

En 1738, Jacques de Vaucanson, artista e inventor, construyó una serie de autómatas para la Academia Francesa de la Ciencia, entre los cuales se hallaba un complejo pato que parecía natural. No solo imitaba los movimientos de un pato real, como mover las alas y comer granos, sino que también podía imitar la digestión. Esta creación puso a los filósofos a pensar: si tenía aspecto de pato y graznaba como un pato, ¿era realmente un pato? Si percibimos que el pato tiene un alma de un tipo diferente, ¿bastaría eso para probar que el pato era consciente de sí mismo, con todo lo que eso implicaba?

El filósofo escocés David Hume rechazó la idea de que el reconocimiento de la existencia era en sí mismo una prueba de conciencia. A diferencia de Descartes, Hume era empirista y desarrolló un nuevo marco científico basado en hechos observables y en argumentos lógicos. Mientras Vaucanson exhibía con orgullo su pato mecánico capaz de digerir (mucho antes de que empezara a hablarse de IA), Hume escribía, en su *Tratado sobre la naturaleza humana*, lo siguiente: «La razón es, y solo debe ser, esclava de las pasiones». En este caso, al hablar de «pasiones», Hume se refiere a «motivaciones no racionales», y apunta a que son los incentivos, y no la lógica abstracta, los que mueven nuestro comportamiento. Si las impresiones son simplemente nuestra percepción de algo que podemos ver, tocar, sentir, gustar y oler, y si las ideas son percepciones de cosas con las cuales no entramos en contacto directo, Hume creía que nuestra existencia y nuestra comprensión del mundo circundante se basaba en un constructo de percepción humana.

A la luz de trabajos más avanzados sobre los autómatas, que se iban volviendo cada vez más realistas, y de reflexiones más profundas sobre los computadores y las máquinas pensantes, el físico y filósofo francés Julien Offray de La Mettrie emprendió un radical —y escandaloso— estudio sobre los humanos, los animales y los autómatas. En un documento escrito en 1747, que publicó inicialmente de manera anónima, La Mettrie sostenía que los humanos son sorprendentemente similares a los animales, y que un simio podría aprender el lenguaje humano, si «se lo entrena adecuadamente». La Mettrie también concluyó que los humanos y los animales son meras máquinas, movidas por el instinto y la experiencia. «El cuerpo humano es una máquina que le da cuerda a su propio mecanismo; [...] el alma no es más que un principio de movimiento, o una parte material y sensible del cerebro.»<sup>12</sup>

La idea de que los humanos son, simplemente, máquinas movidas por la materia (tuercas y engranajes que llevan a cabo una serie de funciones) implicaba que no somos especiales o únicos. También implicaba que tal vez somos programables. Si esto era cierto, y si hasta ese momento habíamos sido capaces de crear patos que parecían reales y monjes en miniatura, de ahí se desprendería que, algún día, los humanos podrían crear réplicas de sí mismos y construir una variedad de máquinas pensantes e inteligentes.

#### ¿ERA POSIBLE CREAR UNA MÁQUINA PENSANTE?

En la década de 1830, matemáticos, ingenieros y científicos empezaron a experimentar, con la esperanza de construir máquinas capaces de hacer los mismos cálculos que los «computadores» humanos. La matemática inglesa Ada Lovelace y el científico Charles Babbage inventaron una máquina llamada Máquina Diferencial y luego postularon una más avanzada, la Máquina Analítica, que usaba una serie de

---

12 «Apes to Androids: Is Man a Machine as La Mettrie Suggests?», <[http://www.charliemccarron.com/man\\_a\\_machine/](http://www.charliemccarron.com/man_a_machine/)>.

pasos predeterminados para resolver problemas matemáticos. Babbage no había concebido que la máquina pudiera hacer nada más allá de calcular números. Fue Lovelace quien, en las notas a pie de página de un documento científico que estaba traduciendo, especuló de manera brillante sobre la posibilidad de hacer una versión más poderosa de la máquina, que podría usarse de otras maneras.<sup>13</sup> Si la máquina podía manipular símbolos, a los cuales se les podían asignar diferentes cosas (como notas musicales, por ejemplo), entonces la máquina podría ser usada para «pensar» por fuera de las matemáticas. Aunque Lovelace no creía que un computador pudiera jamás crear pensamiento original, sí contemplaba la posibilidad de crear un sistema complejo que pudiera seguir instrucciones y, por lo tanto, imitar muchas de las acciones cotidianas de la gente. En aquella época, algunos no le dieron importancia, pero Lovelace había escrito el primer programa computacional completo para una futura máquina muy poderosa, décadas antes de que se inventara la bombilla eléctrica.

Algunos kilómetros más al norte de donde trabajaban Lovelace y Babbage en la Universidad de Cambridge, un joven matemático autodidacta llamado George Boole tuvo una inspiración repentina cuando iba caminando por el campo en Doncaster. Esta idea lo llevó a dedicar su vida a la explicación de la lógica del pensamiento humanos.<sup>14</sup> Su caminata fue el origen de lo que hoy conocemos como «álgebra booleana», que es una manera de simplificar expresiones lógicas (por ejemplo: «y», «o» y «no») usando símbolos y números. Así, al computar «verdad y verdad» obtenemos «verdad», lo que correspondería a interruptores y compuertas físicos en un ordenador. Boole tardó dos décadas en formalizar sus ideas. Tuvieron que pasar otros cien años para que alguien comprendiera que la lógica y la

---

13 Luigi Manabrea, *Sketch of the Analytical Engine Invented by Charles Babbage* (Londres: Richard and John E. Taylor, 1843).

14 Desmond MacHale, *The Life and Work of George Boole: A Prelude to the Digital Age*, nueva ed. (Cork University Press, 2014).

probabilidad booleanas podrían ayudar a los ordenadores a evolucionar y pasar de automatizar las matemáticas básicas a producir máquinas pensantes más complejas. En aquella época no había manera de crear una máquina pensante (pues los procesos, los materiales y la energía todavía no estaban a nuestra disposición), y por eso no fue posible probar la teoría.

El salto de las máquinas pensantes teóricas a los ordenadores que comenzaron a imitar el pensamiento humano se dio en la década de 1930, con la publicación de dos ensayos pioneros: «A Symbolic Analysis of Switching and Relay Circuits», de Claude Shannon, y «On Computable Numbers, with an Application to the *Entscheidungsproblem*», de Alan Turing.

Shannon, que era estudiante de ingeniería eléctrica en el MIT, cursó una materia optativa de filosofía, lo cual constituía una elección peculiar. La obra *Una investigación sobre las leyes del pensamiento*, de Boole, fue la referencia primaria de la tesis de Shannon. Su director de investigación, Vannevar Bush, instó a Shannon a mapear la lógica booleana en circuitos físicos. Bush había construido una versión avanzada de la máquina analítica de Lovelace y Babbage (su prototipo se llamaba Analizador Diferencial), con un diseño peculiar: en esa época no había una teoría sistemática que indicara cómo hacer el diseño de los circuitos eléctricos.

El descubrimiento de Shannon consistió en mapear los circuitos eléctricos de la lógica simbólica booleana y, luego, explicar cómo podía usarse la lógica booleana para crear un circuito funcional que permitiera añadir unos y ceros. Shannon había descubierto que los ordenadores tenían dos niveles: el físico (el contenedor) y el lógico (el código).

Mientras Shannon trabajaba para llevar la lógica booleana a circuitos físicos, Turing ensayaba el traductor del lenguaje universal de Leibniz que pudiera representar todo el conocimiento matemático y científico. Turing se proponía probar lo que llamaban el *Entscheidungsproblem*, o «el problema de las decisiones». El problema, *grosso*

*modo*, podría describirse así: no puede existir un algoritmo que determine si una afirmación matemática arbitraria es verdadera o falsa. La respuesta sería negativa. Turing logró probar que no existe ese algoritmo, pero, paralelamente descubrió un modelo matemático para una máquina computadora de múltiples usos.<sup>15</sup>

Eso lo cambió todo. Turing descubrió que un programa y los datos que este usa podían ser almacenados dentro de un ordenador: de nuevo, esta era una proposición radical en la década de 1930. Hasta ese momento, todo el mundo estaba de acuerdo en afirmar que la máquina, el programa y los datos eran independientes cada uno. Por primera vez, la máquina universal de Turing explicaba por qué los tres estaban unidos entre sí. Desde un punto de vista mecánico, la lógica según la cual operaban los circuitos y los interruptores podía también ser codificada dentro del programa y los datos. Pensemos por un instante en la significación de esta afirmación. El contenedor, el programa y los datos formaban parte de una entidad singular, no muy diferente de los humanos. Nosotros también tenemos un contenedor (nuestro cuerpo), unos programas (las funciones celulares autónomas) y unos datos (el ADN combinado con información sensorial directa e indirecta).

Entre tanto, la larga tradición de los autómatas, que había comenzado 400 años atrás con el monje diminuto que caminaba y oraba, por fin se cruzó en el camino con el trabajo de Turing y Shannon. La compañía estadounidense Westinghouse creó un robot basado en relés y llamado Elektro, para la Feria Mundial Exposición Universal de 1939. Era un gigante burdo, dorado, con ruedas debajo de los pies. Contaba con 48 relés eléctricos que

---

15 El experto en lógica Martin Davis lo explica mejor en *The Universal Computer: The Road from Leibniz to Turing*. «Turing sabía que un algoritmo se especifica, en general, mediante una lista de reglas que una persona puede seguir de una manera mecánica precisa, como una receta en un libro de cocina. Turing pudo demostrar que esa persona podría estar limitada a unas pocas acciones básicas extremadamente simples, sin cambiar el resultado final de la computación. Luego, al demostrar que ninguna máquina que realice solo esas acciones básicas podría determinar si una conclusión propuesta se deriva o no de unas premisas dadas [...] pudo concluir que no existe ningún algoritmo para el *Entscheidungsproblem*».

funcionaban con un sistema de relés telefónicos. Elektro respondía, mediante mensajes pregrabados, a comandos de voz que se enviaban por un auricular de teléfono. Era un ordenador antropomorfizado capaz de tomar decisiones rudimentarias (tales como qué decir) en tiempo real y sin participación directa de un humano.

A juzgar por los titulares de los periódicos, los relatos cortos de ciencia ficción y los noticiarios cinematográficos de la época, salta a la vista que la gente fue tomada por sorpresa: quedó impresionada y desconcertada con estos inventos. El público sentía como si las «máquinas pensantes» ya hubieran llegado, completamente formadas, de la noche a la mañana. En el número de mayo de 1941 de la revista *Astounding Science Fiction*, el escritor de ciencia ficción Isaac Asimov publicó «*Liar!*» [¡Mentiroso!], un relato corto profético. Era una reacción a la investigación que veía desarrollarse en la periferia. En el relato hacía una presentación y una defensa de sus «tres leyes de la robótica»:

1. Un robot no puede hacerle daño a un ser humano ni, a través de la inacción, permitir que un humano reciba daño.
2. Un robot debe obedecer las órdenes dadas a él por los seres humanos, salvo en los casos en que dichas órdenes entren en conflicto con la primera Ley.
3. Un robot debe proteger su propia existencia siempre y cuando dicha protección no entre en conflicto con la primera o la segunda leyes.

Más adelante, Asimov añadió la que él llamo la ley Zeroth, que regía las demás: «Un robot no puede hacerle daño a la humanidad ni, a través de la inacción, permitir que la humanidad reciba daño».