ERIK J. LARSON

EL MITO DE LA INTELIGENCIA ARTIFICIAL

Por qué las máquinas no pueden pensar como nosotros lo hacemos

Shackleton books

EL MITO DE LA INTELIGENCIA ARTIFICIAL

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
—books—

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*

Shackleton
—books—

@Shackletonbooks

www.shackletonbooks.com

Realización editorial: La Letra, S. L.

Diseño de cubierta: Pau Taverna

Conversión a ebook: Iglú ebooks

ISBN: 978-84-1361-202-7

ÍNDICE

Para Brooke y Ben

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 sabemos 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 desarrollado 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 radicalla conversación sobre la IA.

Primera parte
EL MUNDO SIMPLIFICADO

Capítulo 1
El error de la inteligencia

La historia de la inteligencia artificial comienza con las ideas de una persona que contó con una enorme inteligencia humana: el pionero de la informática Alan Turing.

En 1950, Turing publicó un artículo provocador, «Maquinaria computacional e inteligencia», sobre la posibilidad de crear máquinas inteligentes.[1] Fue un texto audaz, que llegó en un momento en el que los ordenadores eran novedosos pero insignificantes, según los parámetros de hoy en día. Aquellas piezas pesadas y lentas de *hardware* servían para acelerar cálculos científicos como el del análisis criptográfico. Tras una larga preparación, se les podían proporcionar fórmulas de física y unas condiciones iniciales, y obtener de forma automática el radio de una explosión nuclear. IBM no tardó en entender su potencial de cara a reemplazar a los seres humanos en sus operaciones comerciales, como la actualización de hojas de cálculo. Pero ver los ordenadores como criaturas «pensantes» requería de cierta imaginación.

La propuesta de Turing se basaba en un entretenimiento popular llamado «el juego de la imitación». En el juego

original, un hombre y una mujer se ocultan a la vista y una tercera persona, el interrogador, les va haciendo preguntas alternativamente. A través de la lectura de sus respuestas tiene que determinar quién es el hombre y quién la mujer. La gracia está en que el hombre tiene que intentar engañar al interrogador, mientras que la mujer se esfuerza por ayudarle, lo cual conduce a que las respuestas de uno y otro lado resulten sospechosas. Turing reemplazó al hombre y a la mujer por un ordenador y una persona. Así nació lo que hoy conocemos como el «test de Turing»: un ordenador y una persona reciben las preguntas mecanografiadas de un juez humano, y si ese juez no logra identificar debidamente quién es el ordenador, el ordenador gana. Turing argumentó que, a partir de ese resultado, no dispondremos de ningún buen motivo para afirmar que la máquina carezca de inteligencia, sin importar que esta sea humana o no. Así, la cuestión de que la máquina disponga de inteligencia reemplazó la cuestión sobre si la máquina puede pensar de verdad.

El test de Turing, en realidad, es muy difícil: ningún ordenador lo ha superado. Por supuesto, en 1950 Turing desconocía este resultado a largo plazo; no obstante, al reemplazar las preguntas filosóficas problemáticas sobre la «consciencia» y el «pensamiento» con un test de resultados observables alentó la visión de la IA como una ciencia legítima con un objetivo bien definido. Mientras la IA cobraba forma durante los años cincuenta, muchos de sus pioneros y seguidores coincidieron con Turing: todo ordenador que pudiera mantener una conversación sostenida y convincente con una persona estaría, tal y como reconoceríamos la mayoría de nosotros, haciendo algo para lo que es necesario el pensamiento (sea eso lo que sea).

LA INTUICIÓN DE TURING / EL INGENIO COMO DISTINCIÓN

Turing se había labrado una reputación como matemático mucho antes de comenzar a escribir sobre IA. En 1936 publicó un artículo corto sobre el significado concreto de la palabra «computador», que en aquel momento se refería a la persona que seguía una serie de pasos para obtener un resultado definido (como la realización de un cálculo).[2] En aquel artículo reemplazó al computador humano por la idea de una máquina que realizara el mismo trabajo. El texto se adentraba en unas matemáticas de gran dificultad. Pero, mientras se refería a las máquinas, no hacía ninguna referencia al pensamiento humano ni a la mente. Las máquinas pueden operar de manera automática, afirmaba Turing, y los problemas que solucionan no requieren de ninguna ayuda «externa» o inteligencia. Esa inteligencia externa —el factor humano— es lo que los matemáticos a veces denominan «intuición».

El trabajo que Turing dedicó en 1936 a las máquinas computadoras ayudó a lanzar la ciencia informática como disciplina, y representó una contribución importante a la lógica matemática. Aun así, al parecer Turing pensó que aquella definición temprana pasaba por alto una cuestión esencial. De hecho, la misma idea de que la mente o las facultades humanas pudieran ayudar a solucionar problemas apareció dos años después en su tesis doctoral, un inteligente pero fallido intento de esquivar uno de los resultados obtenidos por Kurt Gödel, matemático de origen austriaco especializado en lógica (volveremos a él en un rato). La tesis de Turing contiene este curioso pasaje sobre la intuición, que compara con otra capacidad mental a la que llama «ingenio»:

> El razonamiento matemático puede ser considerado, de manera bastante esquemática, como un ejercicio de combinación entre dos facultades, a las que podríamos denominar intuición e ingenio. La actividad de la intuición consiste en realizar juicios espontáneos que no son el resultado de un hilo de razonamientos conscientes. Esos juicios son a menudo correctos, pero de ninguna manera lo son siempre (dejando de lado la cuestión sobre lo que se quiera decir con «correcto»). A menudo resulta posible encontrar otra manera de verificar la corrección de un juicio intuitivo. Por ejemplo, se puede juzgar que todos los números enteros positivos son factorizables en números primos; la argumentación matemática detallada conducirá a idéntico resultado. Esta también incluirá juicios intuitivos, pero serán menos susceptibles a la crítica que el juicio original sobre la factorización. No pretendo explicar esta idea de «intuición» de manera más explícita».

A continuación, Turing pasa a explicar el ingenio:

> En matemáticas, el ejercicio del ingenio consiste en apoyar la intuición a través de una disposición adecuada de las proposiciones, y quizá de las figuras geométricas o de los dibujos. Lo que se pretende es que, cuando estos se encuentren dispuestos de manera verdaderamente correcta, la validez de los pasos intuitivos que sean necesarios no pueda ser motivo de una duda seria.[3]

Aunque su lenguaje se dirija a los especialistas, Turing señala lo evidente: por lo general, los matemáticos escogen sus problemas o «ven» un problema de interés en el que trabajar sirviéndose de una habilidad que cuando menos parece no poder dividirse en pasos, y que, por tanto, no se presta con claridad a la programación informática.

LA PERCEPCIÓN DE GÖDEL

También Gödel pensaba en la inteligencia mecánica. Igual que Turing, estaba obsesionado con la diferencia entre «ingenio» (mecánica) e «intuición» (mente). La distinción que él realizaba era en esencia la misma que la de Turing, aunque con un lenguaje diferente: demostración frente a verdad (o «teoría de la demostración» frente a «teoría de los modelos», en la jerga matemática). ¿Son, en definitiva,

los de demostración y verdad el mismo concepto?, se preguntó Gödel. En caso afirmativo, las matemáticas e incluso la ciencia misma podrían entenderse de manera exclusivamente mecánica. Según esa visión, el pensamiento humano también sería mecánico. El concepto de AI, aunque el término no se hubiera acuñado aún, flotaba sobre la cuestión. ¿Se puede reducir la intuición de la mente, su capacidad para captar la verdad y el significado, a una máquina, a la computación?

Esta era la pregunta de Gödel. Al intentar contestarla, se encontró con un obstáculo que no tardaría en darle fama mundial. En 1931, Gödel publicó dos teoremas de lógica matemática conocidos como los teoremas de incompletitud. En ellos demostró las limitaciones inherentes a todos los sistemas matemáticos formales. Fue un golpe brillante. Gödel demostró de manera inconfundible que las matemáticas —toda la matemática, con ciertas suposiciones directas— no son, hablando en sentido estricto, ni mecánicas ni «formalizables». De manera más específica, Gödel demostró que en todo sistema formal (matemático o informático) han de existir proposiciones Verdaderas, con uve mayúscula, pero que no se pueden comprobar dentro del sistema mismo, sirviéndose de alguna de sus normas. La mente humana puede reconocer esa proposición Verdadera, pero el sistema en el que se ha formulado no la puede demostrar (cosa que sí es demostrable).

¿Cómo alcanzó Gödel esa conclusión? Los detalles son técnicos y complicados, pero la idea básica de Gödel es que podemos tratar un sistema matemático lo bastante complejo para realizar sumas igual que un sistema de significado, casi como si fuera una lengua natural del estilo del inglés o el francés —y lo mismo sirve para todos los sistemas de mayor complejidad—. Al tratarlo de esa manera, posibilitamos que el sistema hable sobre sí mismo.

Y puede contarnos, por ejemplo, que presenta ciertas limitaciones. Esa fue la percepción de Gödel.

Los sistemas formales, como los que aparecen en las matemáticas, permiten la expresión precisa de verdades y falsedades. Por lo general, establecemos lo que es verdad utilizando las herramientas de la demostración —nos servimos de unas reglas para demostrar algo y así sabemos que es indudablemente cierto. Pero ¿hay proposiciones verdaderas que no se puedan demostrar? ¿Puede la mente saber cosas que se le escapen al sistema? En el sencillo caso de la aritmética, expresamos verdades escribiendo ecuaciones como «2 + 2 = 4». Las ecuaciones básicas son proposiciones verdaderas dentro del sistema aritmético, demostrables según las normas de la aritmética. Aquí se da una equivalencia entre lo demostrable y lo verdadero. Antes de Gödel, los matemáticos pensaban que la matemática entera presentaba esa propiedad. Eso implicaba que las máquinas podrían producir en serie todas las verdades de los diferentes sistemas matemáticos limitándose a aplicar las normas de manera correcta. Es una idea hermosa, pero no es cierta.

A Gödel se le ocurrió la extraña pero poderosa propiedad de la autorreferencia. Se puede formar una versión matemática de expresiones autorreferenciales como «Esta proposición no se puede demostrar dentro de este sistema» sin quebrantar las reglas de los sistemas matemáticos. Pero las denominadas «proposiciones autorreferenciales de Gödel» introducen contradicciones en la matemática: si son ciertas, son indemostrables. Si son falsas, puesto que afirman ser indemostrables, en realidad son ciertas. Lo verdadero significa falso, y lo falso, verdadero: es una contradicción.

Retomando el concepto de «intuición», nosotros, los seres humanos, podemos ver que, de hecho, la proposición de Gödel es verdadera, pero, por culpa del resultado de Gödel, sabemos también que las normas del sistema no

pueden demostrarla —en efecto, el sistema se muestra ciego ante aquello que sus normas no alcanzan a cubrir—.[4] Lo que es verdad y lo que es demostrable se desmontan entre sí. Y es posible que pase lo mismo con la mente y la máquina. En cualquier caso, los sistemas puramente formales tienen sus límites. No pueden probar desde su propio lenguaje algo que es cierto. En otras palabras, nosotros podemos ver cosas que al ordenador se le escapan.[5]

El resultado de Gödel representó un duro golpe para la idea, popular en aquel momento, de que todas las matemáticas se podían convertir en operaciones basadas en normas que produjeran una verdad matemática tras otra. El *Zeitgeist* pertenecía al formalismo, no a la conversación sobre las mentes, los espíritus, las almas y demás. En el campo de las matemáticas, el movimiento formalista señaló un giro más amplio de los intelectuales hacia el materialismo científico y, en particular, el positivismo lógico —un movimiento dedicado a erradicar la metafísica tradicional, como el platonismo, con esas formas abstractas que no se podían percibir con los sentidos, y las nociones tradicionales de la religión, como la existencia de Dios—. En efecto, el mundo estaba orientándose hacia la idea de las máquinas de precisión. Y nadie abrazó la causa formalista con tanto vigor como el matemático alemán David Hilbert.

EL DESAFÍO DE HILBERT

A principios del siglo XX (antes de Gödel), David Hilbert había lanzado un desafío al mundo matemático: demostrar que la totalidad de las matemáticas descansaba sobre un fundamento seguro. La ansiedad de Hilbert era

comprensible. Si las normas puramente formales de la matemática no podían demostrar todas y cada una de sus verdades, al menos en teoría era posible que las matemáticas escondieran contradicciones y paparruchas. Que hubiera una contradicción oculta en algún lugar de las matemáticas lo arruinaba todo, porque a partir de una contradicción se puede demostrar cualquier cosa. Y, por tanto, el formalismo ya no servía para nada.

Hilbert expresó el sueño de cualquier formalista: demostrar al fin que las matemáticas eran un sistema cerrado y regido solo por normas. La verdad era tan solo una «demostración». Adquirimos conocimiento cuando nos limitamos a rastrear el «código» de una demostración y confirmamos que no se ha violado ninguna norma. El sueño más amplio de Hilbert, apenas disfrazado, apuntaba en realidad a una cosmovisión, a una imagen del universo en que este mismo fuera un mecanismo. La IA comenzó a cobrar forma como idea, una postura filosófica que también podía demostrarse. El formalismo trataba la inteligencia como si fuera un proceso reglado. Una máquina.

Hilbert lanzó su desafío durante el Segundo Congreso Internacional de Matemáticos, que se celebró en París en 1900. El mundo intelectual le dedicó su atención. El desafío constaba de tres partes principales: demostrar que las matemáticas eran una disciplina completa; demostrar que las matemáticas eran una disciplina consistente y demostrar que las matemáticas eran una disciplina decidible.

Con la publicación de sus teoremas de incompletitud, en 1931, Gödel hirió de muerte las partes primera y segunda del desafío de Hilbert. La cuestión de la decidibilidad quedó sin respuesta. Un sistema es decidible cuando existe un procedimiento definido (una demostración o una secuencia de pasos deterministas y evidentes) para establecer si una proposición construida a partir de las normas de ese sistema es verdadera o falsa. La proposición

2 + 2 = 4 tiene que ser Verdadera, y la proposición 2 + 2 = 5 tiene que ser Falsa. Y sucede lo mismo con todas las proposiciones que se puedan realizar con validez utilizando los símbolos y las reglas del sistema. Puesto que se creía que la aritmética era la base de las matemáticas, demostrar que las matemáticas eran decidibles implicaba demostrar el resultado de la aritmética y sus extensiones. Eso equivaldría a decir que los matemáticos, al «jugar» su partida con reglas y símbolos (la idea formalista), participaban de hecho en un juego válido que nunca conduciría a la contradicción ni al absurdo.

Turing quedó fascinado por el resultado de Gödel, que demostraba no el poder de los sistemas formales, sino más bien sus limitaciones. Se puso a trabajar en la parte que quedaba del desafío de Hilbert y comenzó a pensar en serio si podía existir un proceso de decisión para los sistemas formales. En 1936, con un artículo titulado «Números computables», demostró que no era así. Turing se dio cuenta de que el uso de la autorreferencia por parte de Gödel también podía aplicarse a las preguntas sobre los procesos de decisión, o, en efecto, a los programas informáticos. En especial, se percató de que debían de existir números (reales) que ningún método definido pudiera «calcular» al escribir su expansión decimal, dígito a dígito. Importó un resultado del matemático del siglo XIX Georg Cantor, quien había demostrado que los números reales (aquellos con expansión decimal) eran más numerosos que los enteros, por más que tanto los números reales como los enteros fueran infinitos. Es posible que Turing se subiera sobre hombros de gigantes, pero, al final, su labor en «Números computables» demostró una imposibilidad. Fue un resultado restrictivo: no era posible ningún proceso de decisión universal. En otras palabras, las reglas —incluso en matemáticas— no bastan. Hilbert se había equivocado.[6]

LO QUE IMPLICÓ PARA LA IA

Lo importante de cara a la IA es lo siguiente: Turing refutó que las matemáticas fueran decidibles inventando una máquina, una máquina determinista, que no requería de ninguna intuición o inteligencia para resolver problemas. Hoy en día nos referimos a esa formulación abstracta de una máquina como la máquina de Turing. Ahora mismo estoy tecleando en una de ellas. Las máquinas de Turing son los ordenadores. Que el marco teórico de la informática se implementara como idea colateral, como un medio para obtener un fin diferente, es una de las grandes ironías de la historia intelectual. Mientras trabajaba para refutar que las matemáticas mismas fueran decidibles, Turing fue el primero en inventar algo preciso y mecánico: el ordenador.

En su tesis de 1938, Turing esperaba que los sistemas formales fueran ampliables incluyendo normas adicionales (y a continuación conjuntos de normas, y conjuntos de conjuntos de conjuntos de normas) que pudieran resolver el «problema de Gödel». Descubrió, en cambio, que aquel sistema nuevo y más potente tendría un problema de Gödel nuevo y más complejo. No había manera de sortear la incompletitud de Gödel. No obstante, enterrada bajo las complejidades del razonamiento de Turing sobre los sistemas formales, había una extraña sugerencia que resultaba relevante de cara a la posible existencia de la IA. ¿Y si la facultad de la intuición no se podía reducir a un algoritmo, a las normas de un sistema?

En su tesis de 1938, Turing intentaba encontrar una salida para el resultado restrictivo de Gödel, pero descubrió que era imposible. En su lugar cambió de marcha, se puso a explorar la manera en que, en sus propias palabras, podría «reducir en gran medida» el requisito de la intuición humana a la hora de realizar cálculos. Su tesis tomó en consideración el poder del

ingenio al crear sistemas de normas cada vez más complicados (resultó que el ingenio podía volverse universal: hay máquinas capaces de tomar como referencia a otras máquinas y así dirigir todas las que se puedan construir. Esta percepción, técnicamente una máquina de Turing universal en vez de una simple, iba a convertirse en el ordenador digital). Pero, en su trabajo formal sobre la computación, Turing se había ido de la lengua (quizá de manera involuntaria). Al permitir que la intuición fuera diferente y externa respecto a las operaciones de un sistema puramente formal como es el ordenador, Turing estaba, de hecho, sugiriendo que podían existir diferencias entre los programas de ordenador dedicados a las matemáticas y los matemáticos.

Por tanto, fue curioso el giro que Turing realizó entre sus primeros trabajos de los años treinta y la especulación de amplio espectro acerca de la posible aparición de ordenadores inteligentes en «Maquinaria computacional e inteligencia», que se publicó una década larga después. Hacia 1950, el debate sobre la intuición había desaparecido de los textos de Turing sobre las implicaciones de Gödel. Su interés se trasladó, en efecto, a la posibilidad de que los mismos ordenadores se convirtieran en «máquinas intuitivas». En esencia, decidió que el resultado de Gödel no era aplicable al asunto de la IA: si los seres humanos somos ordenadores muy avanzados, el resultado de Gödel solo implica que hay algunas proposiciones que no podemos comprender o ver como verdaderas, tal y como sucede con otros ordenadores menos complejos. Esas proposiciones podrían ser complejas e interesantes a extremos fantásticos. O tal vez fueran banales pero abrumadoramente complejas. El resultado de Gödel dejaba abierta la cuestión de si la mente no era más que una máquina de gran complejidad, con unas limitaciones muy complejas.

En otras palabras, la intuición había pasado a formar parte de las ideas de Turing acerca de las máquinas y sus poderes. El resultado de Gödel no podía afirmar (según Turing, en cualquier caso) que la mente fuera una máquina o no. Por un lado, la incompletitud sostiene que algunas proposiciones pueden entenderse como verdaderas desde el uso de la intuición, pero que eso no se puede demostrar a partir de un ordenador que se sirva del ingenio. Por el otro, un ordenador más poderoso puede utilizar axiomas (o más bits de código relevante) y demostrar el resultado, mostrando así que la intuición no está lejos de la computación en lo que a este problema se refiere. La cosa se convierte en una carrera armamentística: un ingenio cada vez más poderoso que sustituya a la intuición en problemas cada vez más complejos. Nadie puede anticipar quién ganará la carrera, así que nadie puede argumentar nada —usando el resultado de la incompletitud— sobre las diferencias inherentes entre intuición (la mente) e ingenio (la máquina). Pero, tal y como Turing sin duda sabía, de ser eso cierto, también lo sería al menos la posibilidad de una inteligencia artificial.

Así, entre 1938 y 1950, Turing cambió de opinión acerca del ingenio y la intuición. En 1938, la intuición era el misterioso «poder de selección» que ayudaba a los matemáticos a decidir los sistemas con los que debían trabajar y los problemas que debían resover. La intuición no era algo que se encontrara en el ordenador. Era algo que tomaba decisiones acerca del ordenador. En 1938, Turing no creía que la intuición formara parte de sistema alguno, lo cual sugería no solo que la mente y la máquina eran fundamentalmente diferentes, sino que una IA paralela al pensamiento humano resultaba casi imposible.

Sin embargo, para 1950 había cambiado de parecer. Con el test de Turing, desafió a los expertos e hizo una especie de defensa de la intuición en las máquinas; fue como si preguntara: «¿Por qué no?». Aquello supuso un cambio

radical. Parecía que una nueva visión de la inteligencia comenzaba a cobrar forma.

¿Por qué ese cambio? Entre 1938 y 1950, a Turing le pasó algo ajeno al ámbito de las matemáticas estrictas y la lógica y los sistemas formales. Fue algo que le pasó, de hecho, a toda Gran Bretaña, y ciertamente a la mayor parte del mundo. Lo que pasó fue la segunda guerra mundial.

Capítulo 2
Turing en Bletchley

A Turing le fascinaba el juego del ajedrez —igual que a I. J. «Jack» Good, su colega matemático en tiempos de guerra. Cuando se enfrentaban (solía ganar Good), elaboraban procesos de decisión y reglas de oro para los movimientos ganadores. Jugar al ajedrez implica seguir las reglas del juego (ingenio), pero también parece requerir de cierta percepción (intuición) sobre las jugadas que pueden elegirse según las diferentes posiciones que se den sobre el tablero. Para ganar al ajedrez no basta con aplicar las reglas; en primer lugar, hay que saber qué reglas escoger.

Turing veía el ajedrez como una manera útil (y sin duda entretenida) de pensar sobre las máquinas y la posibilidad de conferirles intuición. Al otro lado del Atlántico, el fundador de la teoría de la información moderna, Claude Shannon, colega y amigo de Turing en Bell Labs, también pensaba en el ajedrez. Más adelante construyó uno de los primeros ordenadores que lo jugaron, una ampliación de la labor que había realizado anteriormente en un protoordenador llamado «el analizador diferencial», que podía convertir ciertos problemas de cálculo en procedimientos mecánicos.[1]

EL PRINCIPIO DE LA SIMPLIFICACIÓN DE LA INTELIGENCIA

El ajedrez fascinaba a Turing y a sus colegas en parte porque parecía que un ordenador podría programarse para jugar sin que la persona que lo programara necesitara saber todo por anticipado. Puesto que los dispositivos informáticos implementaban conectores lógicos como *si-entonces*, *o* e *y*, se podría ejecutar un programa (un conjunto de instrucciones) que generara resultados diferentes dependiendo de los escenarios con los que se encontrara mientras repasaba sus instrucciones. Esa capacidad para cambiar de rumbo según lo que «viera» parecía, a juicio de Turing y sus colegas, simular un aspecto fundamental del pensamiento humano.[2]

Los jugadores de ajedrez —Turing, Good, Shannon y demás— tenían también en la cabeza otro problema matemático con una apuesta mucho más elevada. Trabajaban para sus gobiernos, ayudando a descifrar los códigos secretos que usaba Alemania para coordinar sus ataques contra los barcos comerciales y militares que cruzaban el canal de la Mancha y el océano Atlántico. Turing se comprometió con un esfuerzo desesperado por ayudar a derrotar a la Alemania nazi durante la segunda guerra mundial, y fueron sus ideas sobre computación las que contribuyeron a alterar el curso de la guerra.

BLETCHLEY PARK

Bletchley Park, sita de manera discreta en un pueblo pequeño y alejado del reguero de bombas que caían sobre Londres y la Gran Bretaña metropolitana, era un centro de investigación establecido para ayudar a descubrir la

localización de los *U-boote*, los submarinos alemanes, que causaban estragos en las rutas marinas del canal de la Mancha. Los submarinos nazis representaban un problema capital para las fuerzas aliadas; habían hundido miles de embarcaciones y destruido enormes cantidades de suministros y equipamiento. Para mantener el esfuerzo de guerra, Gran Bretaña necesitaba importaciones de treinta millones de toneladas al año. En un momento dado, los *U-boote* llegaron a reducir esa cantidad en 200.000 toneladas al mes, siguiendo una estrategia de guerra reveladora y potencialmente catastrófica, para la que durante bastante tiempo no hubo réplica. En respuesta, el gobierno británico reunió a un grupo de criptoanalistas, jugadores de ajedrez y matemáticos talentosos para que investigaran la manera de descifrar las comunicaciones con los submarinos, conocidas como «cifrados». (Un «cifrado» es un mensaje oculto. Descifrar un mensaje consiste en convertirlo de nuevo en un texto legible.)[3]

Los códigos se generaban a través de un aparato con aspecto de máquina de escribir conocido como Enigma, que se comercializaba desde los años 1920 pero que los alemanes habían reforzado de manera importante para usarla en la guerra. Las máquinas Enigma modificadas se utilizaron en todo tipo de comunicaciones estratégicas dentro del esfuerzo de guerra nazi. La Luftwaffe, por ejemplo, las usó en su gestión de la guerra aérea, y lo mismo hizo la Kriegsmarine en sus operaciones navales. En general, se consideraba que los mensajes encriptados con la máquina Enigma modificada eran indescifrables.

El papel que Turing desempeñó en Bletchley y su consiguiente ascenso a la categoría de héroe nacional después de la guerra es una historia que ya se ha contado muchas veces. (En 2014, una gran producción cinematográfica, *The Imitation Game [Descifrando Enigma]*, dramatizó su trabajo en Bletchley, así como su rol

consiguiente en el desarrollo de los ordenadores.) El mayor logro de Turing fue relativamente desaborido, según criterios matemáticos puros, porque explotó una vieja idea de la lógica deductiva. El método, al que él y otras personas se referían medio en broma como «turinguismo», se basó en eliminar amplios números de posibles soluciones para los códigos de Enigma encontrando combinaciones en las que hubiera contradicciones. Las combinaciones contradictorias son una imposibilidad; en un sistema lógico no puede darse «A» y «no A» a la vez, tal y como no podemos estar «en la tienda» y «en casa» al mismo tiempo. El turinguismo fue una idea ganadora, y se convirtió en un gran éxito en Bletchley. Logró lo que se había exigido a aquellos «jóvenes genios» recluidos en el laboratorio de ideas al acelerar el descifrado de los mensajes de Enigma. Otros científicos de Bletchley concibieron estrategias diferentes para descifrar los códigos.[4] Sus ideas se ponían a prueba con una máquina llamada Bombe —nombre burlón que provenía de una máquina polaca anterior, la Bomba, y que con toda probabilidad se inspiró en los ruiditos que esta realizaba al terminar cada uno de sus cálculos—. Pensemos en la Bombe como en un protoordenador, capaz de ejecutar diferentes programas.

Más o menos en 1943, el Eje perdió su ventaja bélica en beneficio de las fuerzas aliadas, y ello se debió en no poca medida al esfuerzo continuado de los descifradores de Bletchley. Aquel equipo obtuvo un éxito célebre, y sus miembros se convirtieron en héroes de guerra. Hicieron carrera. Bletchley, mientras tanto, también se reveló como un refugio para el pensamiento dedicado a la computación: la Bombe era una máquina que ejecutaba programas para resolver problemas que los seres humanos por sí mismos no podían solucionar.