



¿Inteligencia mecánica?

La pregunta de Alan Turing

Atocha Aliseda



Aquí se analizarán preguntas sobre la inteligencia mecánica. ¿Cómo surge? ¿Cómo se formula y cómo se responde? ¿Cómo reaccionaron algunos miembros de la comunidad académica ante esta pregunta? Se situarán estas interrogantes en el contexto del nacimiento de las computadoras con Alan Turing como protagonista, pero rastreando sus antecedentes en la noción de cálculo del siglo XVII.

Introducción

¿Máquinas inteligentes? ¿Máquinas que sienten? ¿Recuerda usted *2001: Odisea del espacio*? Esta novela de ciencia ficción, que Arthur Clarke escribe y que Stanley Kubrick lleva a la pantalla, tiene como tripulante de una nave espacial a HAL 9000, una máquina inteligente. Por un error en su programación, HAL se equivoca al anunciar que la nave fallará. La instrucción que reciben desde la Tierra los tripulantes humanos es desconectarla, pero no cuentan con que HAL les lee los labios y se entera de su destino. A partir de este momento, HAL se rebela contra sus compañeros y hace todo lo posible por evitar que la desconecten, lo que la lleva a ocasionar la muerte de todos los tripulantes, menos uno.

HAL no sólo muestra un comportamiento inteligente, sino que también revela sus emociones y reconoce que tiene miedo al ser poco a poco desconectada. Intenta persuadir a David de que no lo haga: “Honestamente pienso que debes sentarte y tomar esto con calma. Tómate una pastilla para el estrés y piensa las cosas de nuevo.” Por si esto fuera poco, la máquina parece estar consciente de sí misma y de sus acciones, cuando expresa lo siguiente: “sé que no todo ha estado bien conmigo”; y cuando finalmente dice: “mi mente se desvanece, lo puedo sentir”.

HAL es el tipo de máquina inteligente que concebía Alan Turing. Nuestro interés es analizar la pregunta relativa a la inteligencia computacional o *inteligencia*



mecánica, como la prefería llamar Alan Turing. ¿Cómo surge? ¿Cómo se formula y cómo se responde? ¿Cómo reaccionaron algunos miembros de la comunidad académica ante esta pregunta de herejes? Situaremos estas preguntas en el contexto del nacimiento de las computadoras, pero rastreando sus antecedentes, en particular en la noción de cálculo del siglo XVII, adjudicada a Gottfried Wilhelm Leibniz (1646-1716).

● **Del cálculo al cómputo: Leibniz y Turing**

La noción de cómputo tiene sus antecedentes en el principio del cálculo, lo que nos remite al siglo XVII, y directamente a Leibniz, uno de los inventores –conjuntamente con Newton– del cálculo diferencial e integral. La notación que construyó para su cálculo lo hizo soñar con la existencia de máquinas que hicieran operaciones para los humanos, dejándolos “libres para el pensamiento creativo”. Su modelo de una máquina que realizaba las cuatro operaciones aritméticas básicas le valió la entrada a la *Royal Society* en Inglaterra. Pero sus sueños fueron mucho más ambiciosos y los intentó hacer realidad con una “idea maravillosa”, como él mismo la llamó: construir un alfabeto del pensamiento donde cada símbolo representara un concepto y que contara con herramientas para la manipulación de símbolos; algo así como un álgebra del pensamiento. Al sistema de caracteres lo llamó *caracteristica universalis*, y al sistema asociado para la manipulación simbólica, *calculus ratiotinator*.

Leibniz estaba muy interesado en la aplicación de su *idea maravillosa* en el terreno legal (tenía estudios de derecho) y su objetivo consistía, nada más y nada menos, que en dirimir disputas. Cada contendiente debía codificar sus opiniones en lo que hoy llamaríamos fórmulas, de modo que la veracidad o falsedad del asunto pudiera establecerse en un sistema mediante aritmética binaria, en respuesta a la orden de *Calculemus!* Imaginemos las implicaciones prácticas y políticas que podría tener un sistema por medio del cual todos los argumentos racionales pudieran validarse de manera concluyente.

Hoy sabemos que no hay un método (de decisión) posible que nos dé una respuesta a la pregunta de si

una fórmula se puede probar. El trabajo de Alan Turing en los años treinta del siglo XX fue clave para establecer los límites de la computación matemática. Demuestra que no hay solución al llamado “problema de la parada”; esto es, no hay un procedimiento de decisión que resuelva si un programa termina o no (para ahondar en este tema, relacionado con la noción de *decidibilidad*, así como para una descripción detallada sobre la noción de máquina universal o máquina de Turing, véase la contribución de Francisco Hernández Quiroz, en este mismo número de *Ciencia*).

“Hay un largo camino desde Leibniz a Turing”; la frase es parte del título de una de las mejores obras (del profesor Martin Davis) sobre el nacimiento y desarrollo de la noción de *computabilidad*. En este largo recorrido, cada capítulo es una parada en uno de los trabajos seminales de otros grandes lógicos que contribuyeron al desarrollo de esta noción: Boole, Frege, Cantor, Hilbert y Gödel.

● **La pregunta sobre la inteligencia: ¿cómo surge?**

Durante el otoño de 1948, Alan Turing se muda a Manchester a trabajar en el Laboratorio Nacional de Física (NPL, por sus siglas en inglés) bajo la tutela de Maxwell Herman Alexander Newman (1897-1984), en un modelo preliminar de una computadora diseñada para experimentos de pequeña escala, y que bautizaron como “el bebé” (Leavitt, 2007). Para probar su funcionamiento y eficiencia, escogieron un problema matemático del siglo XVII –época de Leibniz– sobre números primos.

En 1644, Marin Mersenne (1588-1648) había conjeturado que ciertos números primos toman la forma $2^n - 1$, cuando n es primo. Posteriormente, otros matemáticos propusieron métodos para calcular instancias particulares de esta forma; en especial D. H. Lehmer (1905-1991) encontró en 1932 que $2^{257} - 1$ no es primo. Max Newman programó uno de los métodos diseñados para hacer los cálculos necesarios y “el bebé” fue exitoso en llevarlos a cabo. Así, verificó los hallazgos hechos por humanos.

Por otra parte, la pregunta sobre la *inteligencia mecánica* era ya un tema que circulaba entre algunos investigadores de la época, en varios lugares del mundo.

El estadounidense Norbert Wiener (1894-1964) acababa de publicar su libro *Cybernetics*, y viajó a Inglaterra para entrevistarse con Alan Turing y discutir con él sobre el futuro de las máquinas inteligentes. Los colegas que trabajaban con “el bebé” estaban tan inmersos en su empresa que no se percataron que la noticia sobre su proyecto de investigación había llegado a oídos de especialistas de otras disciplinas, y a la prensa misma. Un neurocirujano de reconocido prestigio, Sir Geoffrey Jefferson (1886-1961), estaba al tanto del proyecto y de la visita previa de Wiener a Inglaterra. El título que eligió para su conferencia magistral (*Lister Oration*) fue justamente “La mente del hombre mecánico”. Al parecer, lo que se proponía era ridiculizar el proyecto encabezado por Max Newman, cuestionándolo de manera tajante, como podemos asentar en sus propias palabras:

No será hasta que una máquina pueda escribir sonetos o componer un concierto, como producto de pensamientos y emociones sentidas –en lugar de ser consecuencia de una combinación azarosa de símbolos–, que podremos estar de acuerdo con que la máquina iguala al cerebro. Esto es, no sólo haber escrito un soneto, sino saber que lo ha escrito (Leavitt, 2007, p. 236; traducción mía).

Al día siguiente, *The Times*, un periódico inglés de amplia circulación, comenta esta conferencia y parafrasea a Jefferson como sigue: “A menos que la computadora pueda crear conceptos y encontrar las palabras para expresarlos por sí misma, no será más inteligente que un perico” (Leavitt, 2007, p. 237). Esta declaración capta la atención de varios lectores, quienes escriben notas al periódico cuestionando la afirmación de que los pericos no son seres inteligentes. Otra reacción a la conferencia de este neurocirujano es una publicación subsecuente en *The Times*, donde se reporta lo siguiente sobre la “mente mecánica” de Newman y su equipo:

Ha realizado [“el bebé”], en cuestión de semanas, un problema cuya naturaleza no podemos revelar ahora, pero sobre el cual podemos decir que se originó en el siglo XVII y que sólo había sido calculado, hasta la fecha, por humanos [...] Además, podemos trazar una analogía entre la estructura de

esta máquina –con todos sus bulbos y fierros– y la del cerebro humano (Leavitt, 2007, p. 237; traducción mía).

Antes de proseguir a analizar la postura de Alan Turing, es interesante notar que desde que se plantea la posibilidad misma de la “inteligencia mecánica”, la discusión sobre el objetivo mismo de este proyecto está presente, a la vez que están en juego distintas concepciones de la inteligencia. Tiempo después, en la década de 1990, el filósofo John Searle (1932-) propone la distinción entre la inteligencia artificial *fuerte* (o robusta) y la *débil*. Los simpatizantes de la primera sostienen que, en principio, es posible construir máquinas inteligentes (como HAL), capaces de entender el lenguaje natural, con la posibilidad de ser conscientes y de tener sentimientos como los humanos (aunque quizá a su propia manera). Los defensores de la inteligencia artificial débil, por su lado, consideran que es posible



simular computacionalmente la cognición humana en ciertos aspectos –sobre todo aquellos que involucran cálculos matemáticos– y no en otros, como los que tienen que ver con la conciencia y las emociones.

Así, visto desde hoy día, pareciera que Jefferson interpreta el proyecto del laboratorio en Manchester, donde Turing trabajaba, como uno de inteligencia artificial fuerte, y se opone rotundamente a él; mientras que Newman parece abocarse a la versión débil de la inteligencia artificial, pues sólo defiende que su computadora “el bebé” es capaz de resolver problemas matemáticos, aquellos que requieren una gran capacidad de cálculo, pero no menciona nada con respecto a si la máquina pueda estar consciente de lo que está haciendo o si es capaz de sentir.

¿Cuál era la postura de Alan Turing?

En mi opinión, no hay una respuesta fácil ni definitiva a esta pregunta. Por un lado, Turing concibe máquinas con la capacidad de sentir, lo que visto desde hoy día lo sitúa como promotor de una inteligencia

Turing concibe máquinas con la capacidad de sentir, lo que visto desde hoy día lo sitúa como promotor de una inteligencia artificial fuerte; sin embargo, su propuesta matemática y computacional concibe al pensamiento como cálculo de tareas matemáticas, lo que sugiere que aboga más bien por una inteligencia artificial débil

artificial fuerte. Por otro lado, su propuesta matemática y computacional concibe al pensamiento como cálculo de tareas matemáticas, lo que sugiere que aboga más bien por una inteligencia artificial débil. Sin embargo, su postura no es contradictoria, pues podríamos interpretar que si bien su aspiración es por la inteligencia artificial fuerte, el trabajo que desarrolla se enfoca a la débil, aunque bien podría extenderse a uno que busque la inteligencia artificial fuerte. De hecho, éste ha sido el argumento del proyecto de la inteligencia artificial desde su inicio: llegará el momento en que las máquinas sean capaces de pensar y se comporten como nosotros, pero aún no tenemos los recursos –ni de *hardware* ni de *software*– para lograrlo.

En cualquier caso Turing, en seguimiento al debate sostenido en *The Times*, monta en cólera y hace gala de un humor irónico, con la siguiente respuesta:

Esto es sólo un adelanto de lo que está por venir, y es sólo la sombra de lo que será. Debemos tener alguna experiencia con la máquina antes de conocer realmente sus capacidades. Pasarán quizá muchos años antes de que nos acostumbremos a estas nuevas posibilidades, pero no veo por qué no debemos entrar en todos los campos que han sido cubiertos normalmente por el intelecto humano, y eventualmente competir en igualdad de condiciones. No creo ni siquiera que podamos pintar una línea divisoria en la escritura de sonetos, aunque de entrada esta comparación me parece un poco injusta, porque los sonetos escritos por una máquina serán mejor apreciados por otra máquina (Leavitt, 2007, p. 237-238; traducción mía).

Ese mismo año, 1948, Turing se aboca a defender su proyecto y da una charla en Manchester titulada “Maquinaria inteligente: una teoría de herejes”, que comienza como sigue: “No puedes construir una máquina que piense por ti. Esto es un lugar común que usualmente se acepta sin cuestionamiento alguno. El objetivo de este trabajo será cuestionarlo.” Y continúa: “Sostengo que se pueden construir máquinas que simularán el comportamiento de la mente humana de manera muy aproximada” (Turing, 1948, p. 472; traducción mía).

En esta charla, Turing expone de manera muy clara –aunque general– su concepción de una “máquina in-

teligente”. La idea central es el fundamento de lo que hoy conocemos como *aprendizaje por computadora* (para ahondar en este tema, véase la contribución de Eduardo Morales en este mismo número de *Ciencia*).

Inicialmente, dice: una máquina es como un bebé al que hay que enseñar diversas tareas. A través de la educación, este crío puede convertirse en un infante, y eventualmente en un adulto. El aprendizaje se da por medio de la *experiencia*; esto es, gracias a la interacción con un instructor humano, quien no sólo le enseña tareas matemáticas y de otra índole intelectual, sino que también expone a la máquina a sensaciones. Por ejemplo, el instructor le enseña a reconocer los estados de dolor y de placer, de tal manera que, basada en experiencias anteriores, la máquina eventualmente distingue las experiencias dolorosas de las placenteras.

Es clave que la máquina cuente con una memoria que le permitirá almacenar experiencias pasadas y usarlas para procesar las actuales. La máquina, sin embargo, no será infalible: cometerá errores, mismos que le serán señalados por el instructor —a través de castigos—, lo que formará parte de su educación, al parecer de carácter conductista.

Hay, además, otros dos elementos interesantes que Turing vislumbra en cuanto al comportamiento de estas máquinas: el azar y el poder sobre los humanos. Por un lado, su comportamiento no estará completamente determinado de antemano; y, por el otro, de acuerdo con Turing: “Parece probable que una vez que el método del pensamiento mecánico comience, no pasará mucho tiempo en que supere a nuestros pobres poderes” (Turing, 1948, p. 475; traducción mía).

¿Pueden pensar las máquinas?

Finalmente, llegamos al punto que usualmente es el de partida cuando se aborda la pregunta sobre la inteligencia mecánica. En 1950, Alan Turing publica un artículo en la revista de filosofía *Mind*, titulado “Maquinaria computacional e inteligencia” (*Computing machinery and intelligence*), mismo que bien podría haberse llamado “¿Pueden pensar las máquinas?”, ya que lo que propone es, por un lado, situar la pregunta sobre la inteligencia mecánica en una versión del juego de imitación, lo que se conoce ahora como la *Prueba de*



Turing (véase el artículo de Carlos Gershenson, en este mismo número de *Ciencia*). Por otro lado, su propósito es convertir esta pregunta en una pregunta computacional lo más precisa posible, haciendo uso de lo que ahora conocemos como *Máquina de Turing*. Es sus propias palabras:

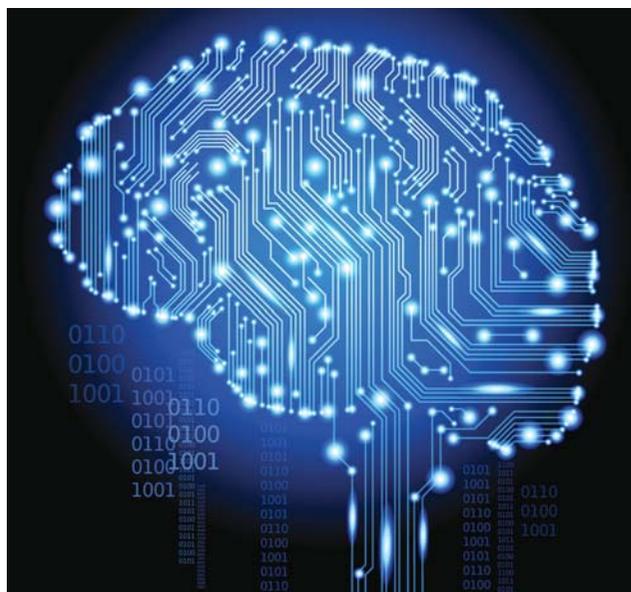
Se sugirió tentativamente que la pregunta ¿pueden pensar las máquinas?, debiera ser sustituida por esta otra: ¿hay computadoras digitales imaginables, que se desempeñarían bien en el juego de imitación? Si quisiéramos, podríamos hacer esta afirmación someramente más general, y preguntar: ¿hay máquinas de estado finito que se desempeñarían bien? Dada la propiedad de universalidad [de las máquinas de Turing], vemos que cualesquiera de estas preguntas es equivalente a esta otra: Fijemos nuestra atención en cierta máquina digital C. ¿Es verdad que si modificamos esta computadora de tal forma que tenga suficiente almacenamiento e incrementamos su velocidad de acción de manera adecuada y la dotamos del programa adecuado, C puede jugar el papel de la parte de A [una mujer] en el juego de imitación, siendo la parte B un hombre? (Luger, 1995, p. 31; traducción mía).

Turing formula así la pregunta sobre la inteligencia mecánica y propone una prueba de simulación de la misma en este juego de imitación, representado por una máquina de Turing. Este artículo, publicado en *Mind*, es precisamente el texto más conocido de Alan Turing

sobre este tema, y se considera el trabajo que sienta los fundamentos del proyecto de la inteligencia artificial y las ciencias cognitivas (para un recuento del inicio de la inteligencia artificial recomendamos el artículo de Carlos Gershenson, en este mismo número de *Ciencia*, y el texto de Aliseda, 2007, sobre la emergencia de las ciencias cognitivas). En este texto, Turing se propone también responder a algunas objeciones muy fuertes a su pregunta, de las cuales resaltamos la afirmación de que una máquina no puede ser el sujeto de su propio pensamiento, una forma de conciencia. Turing da un ejemplo numérico de esta posibilidad y va más allá cuando afirma: “Observando los resultados de su propio comportamiento, la máquina puede modificar sus programas para lograr algún propósito de manera más eficaz. Éstas son posibilidades en un futuro cercano, y no sueños utópicos” (Luger, 1995, p. 37; traducción mía).

Impacto en la filosofía: la mente como computadora

El modelo de la computadora propuesto por Alan Turing tuvo un gran impacto en la filosofía, en particular en la de la mente. El artículo de *Mind* se considera un pilar, y la analogía “la mente es como una computadora” sirvió de puente entre la filosofía de la mente y la inteligencia artificial. La primera propor-



cionó la base conceptual, y la segunda las herramientas para representar y manipular el conocimiento.

Una de las tesis fundamentales que subyace a esta analogía es la siguiente: *el pensamiento es el resultado de representaciones mentales y procesos computacionales que operan sobre estas representaciones*. Dicha tesis asume que la mente posee representaciones análogas a las estructuras de datos, y procedimientos computacionales análogos a los algoritmos. La *Teoría computacional de la mente* combina una teoría de representaciones mentales con un enfoque computacional del razonamiento. Este enfoque, conocido como funcionalismo, es una propuesta original del filósofo Hilary Putnam, y es desarrollada posteriormente y a fondo por Jerry Fodor, quien pretende dar cuenta del lenguaje del pensamiento.

Sin embargo, no hay un único modelo de representaciones de la mente, como tampoco hay un enfoque computacional único. La propuesta funcionalista, por ejemplo, asume que la mente es una máquina de Turing que opera sobre estructuras en forma de reglas de transición entre estados. En esta visión, los estados mentales se corresponden con estados físicos del cerebro; lo que cuenta es lo que las funciones puedan realizar, independientemente del *hardware* que les dé cuerpo. Así, podemos decir que un robot tiene ciertos estados mentales, expresados en *chips* de silicón en lugar de neuronas.

Por otra parte, el enfoque conexionista ha propuesto ideas novedosas de representación y computación que usan una representación de neuronas y sus conexiones como inspiraciones para dar cuerpo a las estructuras de datos, así como las activaciones neuronales como inspiraciones para modelar algoritmos. En el medio computacional, este enfoque se conoce como *redes neuronales*, y tiene también antecedentes en el trabajo de Turing (para saber más del tema, véase la contribución de Pedro Miramontes en este mismo número de *Ciencia*).

A manera de conclusión

La pregunta sobre la *inteligencia mecánica* que se hizo Alan Mathison Turing la respondió haciendo uso de su modelo formal de máquina universal, y la situó en el juego de imitación, con el que pretende ofrecer una prueba de simulación de la inteligencia.

La formulación de esta pregunta –a la manera de Turing– es, sobre todo, una contribución conceptual. Es un paso imprescindible para el entendimiento –preciso y formal– de la inteligencia computacional, y con ello da fundamento a la inteligencia artificial.

Sin embargo, la contribución que me parece aún más importante y de más largo alcance es que la pregunta sobre la inteligencia mecánica que se hace Alan Turing apunta directamente a la pregunta sobre la inteligencia humana, pues en la medida en que podamos caracterizar la segunda, podremos ubicar mejor a la primera. A partir de la pregunta de Alan Turing, la investigación sobre la inteligencia humana en diversas áreas del conocimiento cambió de manera radical. Analizar esta cuestión tan fascinante, sin embargo, nos llevaría fuera de los límites de esta nota.

Atocha Aliseda es investigadora del Instituto de Investigaciones Filosóficas de la UNAM, miembro del Sistema Nacional de Investigadores y de la Academia Mexicana de Ciencias. Obtuvo el doctorado en Filosofía y Sistemas Simbólicos por la Universidad de Stanford. Es especialista en lógica, está interesada en la metodología de las ciencias y en las lógicas del descubrimiento científico. Su tema actual de investigación se centra en las nociones de hipótesis y evidencia, en particular en el diagnóstico médico, con las lógicas adaptativas como herramienta formal.
www.filosoficas.unam.mx/~atocha/home.html
atocha@filosoficas.unam.mx

Para saber más

- Davis, M. (2012), *The universal computer: the road from Leibniz to Turing*, CRC Press.
 Fodor, J. (1975), *The language of thought*, Nueva York, Crowell.
 Wiener, N. (1948), *Cybernetics*, Nueva York, Wiley.

Bibliografía sobre Alan Turing

- El lector interesado puede consultar la bibliografía de Alan Turing compilada por su biógrafo Andrew Hodges en: www.turing.org.uk/sources/biblio.html.
 Página electrónica sobre primos de Mersenne: <http://primes.utm.edu/mersenne/>.

Referencias específicas

- Aliseda, A. (2007), “Emerge una nueva disciplina: las ciencias cognitivas”, *Ciencias*, Facultad de Ciencias, UNAM, 88, 22-31.
 Copeland, B. J. (2004), *The essential Turing: seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life. Plus The secrets of Enigma*, Oxford University Press.
 Leavitt, D. (2007), *The man who knew too much. Alan Turing and the invention of the computer*, Phoenix.
 Luger, G. F. (editor, 1995), *Computation and Intelligence*, American Association for Artificial Intelligence/MIT Press.
 Searle, J. R. (1980), “Minds, brains, and programs”, *Behavioral and brain sciences*, 3(3):417-457.
 Turing, A. M. (1948), “Intelligent machinery: a heretical theory”, conferencia impartida en Manchester, transcripción en The Turing Archive, incluido en Copeland, 2004.
 Turing, A. M. (1950), “Computing machinery and intelligence”, *Mind*, 59(236):433-460.