

Capítulo 9

¿Pueden pensar las máquinas?

Hubo un tiempo en que tuvo que parecer sumamente improbable que las máquinas pudieran aprender a dar cuenta de sus necesidades mediante sonidos, aún yendo éstos dirigidos a oídos humanos. ¿No será lícito, entonces, concebir que pueda llegar el día en que ya no sean necesarios nuestros oídos, sino que la audición se produzca gracias a la delicada construcción de la máquina, el día en que su lenguaje haya trascendido del grito animal a un discurso tan intrincado como el nuestro?

Samuel Butler, Erewhon

Alan Mathison Turing, matemático inglés fallecido en 1954 cuando sólo contaba 42 años, ha sido, entre los pioneros de las ciencias del cómputo, uno de los más creativos. En nuestros días se le conoce sobre todo por la idea de una máquina hipotética, llamada «máquina de Turing». Echaremos aquí una rápida ojeada a estas máquinas y nos detendremos luego en una de las ideas menos conocidas de Turing, el juego de Turing, que conduce a profundas controversias de carácter filosófico, hoy todavía por resolver.

Una máquina de Turing es una «caja negra» (una máquina cuyo mecanismo no se especifica) capaz de ir inspeccionando una cinta ilimitada dividida en casillas. La caja puede tomar un número finito cualquiera de estados. En la cinta hay una porción finita cuyas casillas no están en blanco; cada una de éstas porta un único símbolo tomado de entre una colección finita prefijada. Al inspeccionar una casilla, la caja puede dejar intacto el símbolo que contenga; puede borrarlo; puede borrarlo e imprimir en su lugar otro símbolo; o puede imprimir un símbolo en una casilla vacía. La cinta puede entonces desplazarse una casilla hacia la derecha o la izquierda, o permanecer quieta; por su parte, la caja puede persistir en su estado o saltar a un estado diferente.

La conducta de la máquina en cada una de las combinaciones de símbolo y estado queda determinada por una tabla de reglas. La tabla define totalmente la máquina de Turing

concreta de que se trate. Existe una infinidad numerable (es decir, de cardinal aleph-sub-cero) de posibles máquinas de Turing, cada una diseñada para una tarea específica; y la estructura de la máquina puede diferir mucho en sus símbolos, estados y reglas, según la tarea a ejecutar.

Un buen procedimiento para captar la esencia de las máquinas de Turing consiste en construir una, aunque sea trivial [véase la Figura 47].



Figura 47. Máquina de Turing capaz de efectuar adiciones

En la cinta de papel vemos ocho casillas marcadas 1111 + 111, que denotan la suma de 4 más 3 en el sistema «unario», donde para expresar el entero n se escriben n palotes, n «unos», por ejemplo. Para construir la máquina recortaremos en cartulina un cuadrado no muy grande (la caja negra) y en él dos rendijas, por donde se hará deslizar la cinta como se muestra en la figura. Se ajusta la cinta de forma que sea visible el primer 1. La tabla de instrucciones que acompaña la ilustración enumera todas las instrucciones necesarias. Empecemos suponiendo que la máquina se encuentra en el estado A. Consultamos en la tabla la instrucción correspondiente al símbolo 1 y el estado A, y ejecutamos lo que dice: borrar el 1, desplazar la cinta un cuadro hacia la izquierda (para poder explorar la casilla adyacente a la derecha) y suponer que la máquina ha adoptado el estado B. Se prosigue de esta forma hasta que la tabla ordene detenernos.

Siguiendo correctamente las instrucciones, la máquina borrará el primer 1, y va desplazando la cinta hacia la izquierda, casilla por casilla, hasta alcanzar el signo «más». Una vez alcanzado, cambiará el + por un 1, y se detendrá. El contenido de la cinta será entonces 1111111, es decir, 7. Como es evidente, las sencillas reglas anteriores dejan programado el dispositivo para sumar cualquier par de números enteros en notación unaria, por grandes que sean.

No cabe duda de que como procedimiento de sumación este método es bien fastidioso; pero debemos recordar que el objetivo de Turing era reducir el cálculo mecánico a un esquema abstracto sencillo, que facilitara así el análisis de toda clase de espinosos problemas teóricos, como, por ejemplo, qué puede ser computado y qué no. Turing demostró que su dispositivo ideal puede ser programado para realizar, en su desmañado estilo, cualquier cosa que pueda ejecutar el más potente ordenador electrónico. Lo mismo que cualquier ordenador y que el cerebro humano la máquina de Turing está limitada por el hecho de que ciertos cálculos (como los necesarios para hallar el valor de «pi») exigen infinito número de pasos, y por otro lado, debido a que ciertos problemas son intrínsecamente insolubles, es decir, se sabe que no puede existir ningún algoritmo, ningún proceso perfectamente detallado, que permita resolverlos. Una «máquina universal de Turing» es capaz de llevar a cabo cualquier tarea que pueda efectuar una máquina de Turing especialmente concebida para esa tarea. En breve, la máquina universal es capaz de computar todo lo que sea computable.

En 1950, la revista inglesa *Mind*, dedicada a temas filosóficos, publicó un artículo de Turing, «Computing Machinery and Intelligence». Desde aquella fecha, el artículo ha sido recogido en diversas antologías, entre ellas, en *The World of Mathematics* de James R. Newman (hay traducción española: *Sigma, El mundo de las matemáticas*, Ed. Grijalbo). Allí, Turing empezaba diciendo: «Me propongo examinar la cuestión ¿Pueden pensar las máquinas?». Así planteada, decía Turing, la pregunta era demasiado vaga para poder darle alguna respuesta significativa. Turing proponía entonces otra cuestión mucho más restringida, relacionada con ésta: ¿Es posible enseñar a un ordenador a ganar el «juego de imitación», hoy comúnmente conocido por juego de Turing o test de Turing?

Turing inspiró su test en un juego de salón. Un hombre y una mujer se encierran en distintas habitaciones. Un interrogador, da igual hombre que mujer, va haciéndoles preguntas a los jugadores. Las preguntas son formuladas a través de un intermediario; el correveidile trae las respuestas, de vuelta, escritas a máquina. Cada jugador se propone convencer al preguntón que él o ella es, en realidad, la mujer, pongamos por ejemplo. El interrogador gana el juego cuando atina quien está diciendo la verdad.

Supongamos, decía Turing, que uno de los jugadores sea sustituido por una máquina capaz de aprender, a la que hemos enseñado a conversar en un lenguaje natural. ¿Es posible que una máquina así logre engañar al inquiridor, si tanto la máquina como su compañero humano se esforzasen al máximo en convencer al interrogador de que él, ella o ello son verdaderamente humanos?

El significado de «engañar» queda desdibujado por varias imprecisiones. ¿Cuánto tiempo puede durar el interrogatorio? ¿Cuán inteligente es el interrogador? ¿Cuán inteligente es la persona que compite con la máquina? Un ordenador moderno podría superar el test de Turing si el interrogador fuese un niño que tan sólo pudiera formular unas cuantas preguntas. Es verosímil que no se produzcan en este campo avances espectaculares, como probablemente tampoco se produjeron en la evolución del intelecto humano. Las máquinas conversadoras podrían ir mejorando gradualmente, resistiendo diálogos más y más largos frente a interrogadores cada vez más perspicaces. Quizá llegue un día en que tan sólo un potentísimo ordenador electrónico sea capaz de discriminar sistemática y acertadamente las personas de las máquinas. El propio Turing hizo una predicción cautelosa. Hacia el año 2000, escribió, los ordenadores tendrán la facilidad de palabra suficiente como para despistar a un «interrogador corriente» alrededor del 30 por 100 de las veces al cabo de «unos cinco minutos» de conversación. Turing imaginó un diálogo que pudiera ser típico:

P. Vamos a ver, componga, un soneto dedicado a sus padres.

R. No cuente conmigo para eso. Nunca he sabido escribir poesía.

P. Sume 34.957 y 70.764.

R. (Pausa de unos 30 segundos antes de contestar). 105.621.

P. ¿Sabe jugar al ajedrez?

R. Sí.

P. Tengo mi rey en 1R. No tengo más piezas. Usted tiene su rey en 6R, una torre en 1T, y nada más. Juega usted. ¿Qué debe hacer?

R. (Tras pausa de unos 15 segundos). T8T. Mate.

Existen ya muchos programas que permiten que un ordenador simule conversaciones en lenguaje natural. Joseph Weizenbaurn, ingeniero eléctrico del M.I.T., ha preparado uno que él ha bautizado «Eliza» en recuerdo de la Eliza Doolittle del *Pygmalion* de Shaw. El *New York Times* (15 de junio de 1968) reseñaba el siguiente diálogo mecanográfico entre una joven y un ordenador I.B.M. 7094, que suplantaba a un hipotético psicoanalista:

Mujer: Todos los hombres son iguales.

Ordenador: ¿En qué aspecto?

- M. En que siempre están dándonos la lata con una cosa u otra.
- O. ¿Puede darme algún ejemplo concreto?
- M. Bueno, para que vea, mi novio me ha hecho venir aquí.
- O. ¿Su novio la ha hecho venir aquí?
- M. Sí. Dice que estoy siempre deprimida.
- O. Lamento mucho que esté usted deprimida.
- M. Pero es cierto. No soy feliz.
- O. ¿Cree usted que venir aquí le servirá de ayuda?
- M. Bueno, lo que está claro es que necesito que me ayuden.
- O. ¿De qué cree usted que le serviría que le ayudasen?
- M. Quizás... para llevarme mejor con mi madre.
- O. Hábleme de su familia.
- M. Mi madre se ocupa mucho de mí.
- O. ¿Hay alguien más en su familia que se ocupe de usted?
- M. También mi padre.
- O. ¿Su padre?
- M. Hay cosas en las que usted se parece a mi padre.
- O. ¿Qué parecido me encuentra usted?
- M. Usted no es nada agresivo, pero tengo la impresión de que no quiere que yo me dé cuenta.
- O. ¿Por qué piensa usted que no soy agresivo?
- M. Porque nunca me rebate lo que digo.

El diálogo prosigue en la misma tónica y no difiere gran cosa de la conversación entre un paciente y un terapeuta no directivo. El programa no era tan siquiera un programa «autodidacta». Weizenbaum admitió de buena gana que el ordenador no «comprendía» nada de lo que allí se decía. Desde luego, no podría superar el test de Turing. Supongamos, empero, que hacia el año 2000 haya ordenadores capaces de afrontar el juego de Turing con tanto éxito como ahora son capaces de jugar a las damas o al ajedrez. ¿Qué revelaría eso si es que revela algo acerca de la naturaleza de la «mente» de la máquina?

Los lectores de la famosa novela *2001, una odisea del espacio*, de Arthur C. Clarke, recordarán que en ella se dice que HAL, ordenador parlante de la nave espacial, «piensa», pues es capaz de «superar fácilmente el test de Turing». (Las siglas HAL provienen de computador *heurísticamente* programado y *algorítmico*; pero seguramente, cuando Clarke eligió este nombre se proponía hacerle al lector un juego de palabras más ingenioso. ¿Sabrá

descubrirlo usted?) ¿Podemos decir que HAL piensa, o tan sólo remeda al pensamiento? Turing opina que, llegado el momento en que la habilidad conversatoria de los ordenadores les permitieran superar su test, nadie dudaría en admitir que son capaces de pensar.

Inmediatamente surgen docenas de cuestiones enormemente embrolladas. ¿Podría sufrir la timidez un ordenador semejante? ¿Podría experimentar emociones? ¿Y tener sentido del humor? En pocas palabras, ¿tendríamos que considerarlo «persona»? ¿O tan sólo es una máquina muerta, que ha sido construida para imitar la conducta de las personas?

Recordemos que L. Frank Baum ya escribió sobre el robot TikTok, un robot que «piensa, habla, actúa y hace todo, excepto vivir».

Qué duda cabe, que si un ordenador saliera triunfante de los tests de Turing, lo único que se habría demostrado es que los ordenadores pueden imitar el discurso de los humanos con perfección suficiente como para superar tales pruebas. Pensemos por un momento que alguien, en la Edad Media, hubiera pensado en someter a los tulipanes a la siguiente prueba de autenticidad: ¿Será posible producir un tulipán de orfebrería, tan perfecto que a simple vista sea indistinguible de los del jardín? Hoy se fabrican tulipanes artificiales capaces de superar esta prueba de «autenticidad». Pero eso nada nos dice sobre la capacidad de los químicos para sintetizar compuestos orgánicos, ni nos garantiza que nadie sepa construir un tulipán capaz de crecer como los tulipanes de jardín. Al igual que hoy nos ocurre que al tocar lo que pensamos que es una flor exclamamos con sorpresa « ¡Anda, si es artificial!, no parece en absoluto impensable que mañana, tras sostener una larga conversación con lo que pensamos que es una persona, al abrir una puerta descubramos atónitos que habíamos estado charlando con un ordenador.

Keith Gunderson, en un importante artículo de 1.964 donde criticaba a Turing por haber cargado excesivamente las tintas en la importancia de su test, expresaba así su punto de vista: «Al fin y al cabo, la perforadora de vapor pudo más que John Henry en la tarea de excavar túneles de ferrocarril. Pero eso no demostró que la máquina perforadora tuviera músculos; por el contrario, demostró que para excavar túneles de ferrocarril no se precisa de músculos.»

El test de Turing experimentó un giro curioso durante una conferencia de Michael Scriven, más tarde recogida con el título «The Compleat Robot: A Prolegomena to Androidology», en *Dimensions of Mind*, recopilado por Sidney Hook. Scriven concedía que la habilidad dialéctica no demostraba que el ordenador poseyera otros atributos de la «persona». Supongamos, empero, que uno de estos computadores parlantes llegase a aprender el significado de la palabra «verdad» (por ejemplo, en el sentido de correspondencia que Alfred Tarski ha definido con precisión) y que a continuación se le programa para que nunca pueda mentir.

«De esta forma, el robot queda incapacitado para ejercer de ayuda de cámara, de redactor publicitario o de político, pero en cambio puede prestarnos ahora nuevos servicios.» Ahora podemos preguntarle si tiene conciencia de existir, si tiene emociones, si ciertos chistes le parecen graciosos, si actúa por propia voluntad, si le gusta la poesía de Keats. Y otras semejantes, en la confianza de que nos va a dar respuestas correctas.

Cabe la posibilidad de que la «máquina de Scriven» (como la ha bautizado uno de los varios filósofos que en otros capítulos de la antología de Hook comentan el artículo de Scriven) respondiera negativamente a todo lo anterior. Pero si diera respuestas afirmativas, arguye Scriven, tendríamos tanta justificación para creerlo como para creer a un ser humano, y ninguna para no admitirlo como «persona».

No existe acuerdo entre los filósofos con respecto a los razonamientos de Turing y Scriven. En una nota breve titulada «The Supercomputer as Liar», Scriven replicaba a algunos de sus críticos. Por otra parte, Mortimer J. Adler, en su libro *The Difference of Man and the Difference It Makes*, considera que el criterio de Turing es «cosa de todo o nada», y que el éxito y el fracaso en construir computadoras capaces de superarlo servirán, respectivamente, para debilitar o reforzar la creencia de que el hombre es radicalmente diferente de cualquier máquina o animal infrahumano.

La existencia de máquinas capaces de dialogar, ¿lograría verdaderamente cambiar las creencias de los humanos con respecto al carácter singular de su naturaleza? No cuesta demasiado imaginar uno de nuestros programas humorísticos de televisión dentro de 50 años. Los invitados al programa improvisan chistes y juegos de palabras en compañía de un robot presentador en cuya memoria se han atiborrado uno o dos millones de chascarrillos, y al que humoristas humanos han logrado infundir el arte de la pausa en la conversación intencionada. Por mi parte, dudo mucho que nadie admitiera que el robot «tiene sentido del humor», como ningún jugador derrotado por un autómatas ajedrecístico estaría dispuesto a admitir que se ha enfrentado a una máquina de naturaleza radicalmente distinta a la del autómatas que juega al «tres en raya». Después de todo, las reglas de la semántica y sintaxis no son tan radicalmente diferentes de las del ajedrez.

En cualquier caso, el debate continúa, embrollado por prejuicios metafísicos y religiosos además de problemas lingüísticos muy complejos. Todos los clásicos enigmas sobre el cuerpo y el alma, sobre la naturaleza de la persona, vuelven a ser de actualidad, sólo que ahora planteados con nueva terminología. Resulta difícil predecir qué principios sobrevivirán y cuáles serán pisoteados, ni cómo, al ir perfeccionándose los robots del futuro, se verán influidas las cuestiones filosóficas fundamentales hoy en plena controversia.

Hace 100 años, cuando Samuel Butler explicaba en *Erewhon* por qué los erewhonianos decidieron destruir sus máquinas, temerosos de verlas convertidas de siervas en señoras, la advertencia de Butler fue considerada pura sátira, traída por los pelos. Hoy, la lectura de esos mismos capítulos impresiona por su carácter profético. «Las máquinas tienen hoy un nivel de conciencia muy bajo», escribía Butler, «mas no por eso tenemos seguridad alguna de cual será el definitivo desarrollo de la conciencia mecánica. Un molusco tampoco ofrece un elevado nivel de conciencia. Reflexionemos, empero, sobre el extraordinario desarrollo de las máquinas en estos últimos siglos, y cuán lentamente, en cambio, están avanzando los reinos animal y vegetal. En comparación con el tiempo pasado, las máquinas de más alta organización no son, por así decirlo, cosa de ayer, sino de los cinco últimos minutos».

Soluciones

Si tomamos en el alfabeto la siguiente a cada una de las letras que componen el nombre HAL resulta IBM. En el filme, el logotipo de IBM es visible en los terminales de visualización de HAL, y todo el mundo supuso que Clarke había desplazado las letras intencionadamente. Por su parte, Clarke me ha asegurado que tal hecho es completamente accidental, y que él fue el primer sorprendido al enterarse.