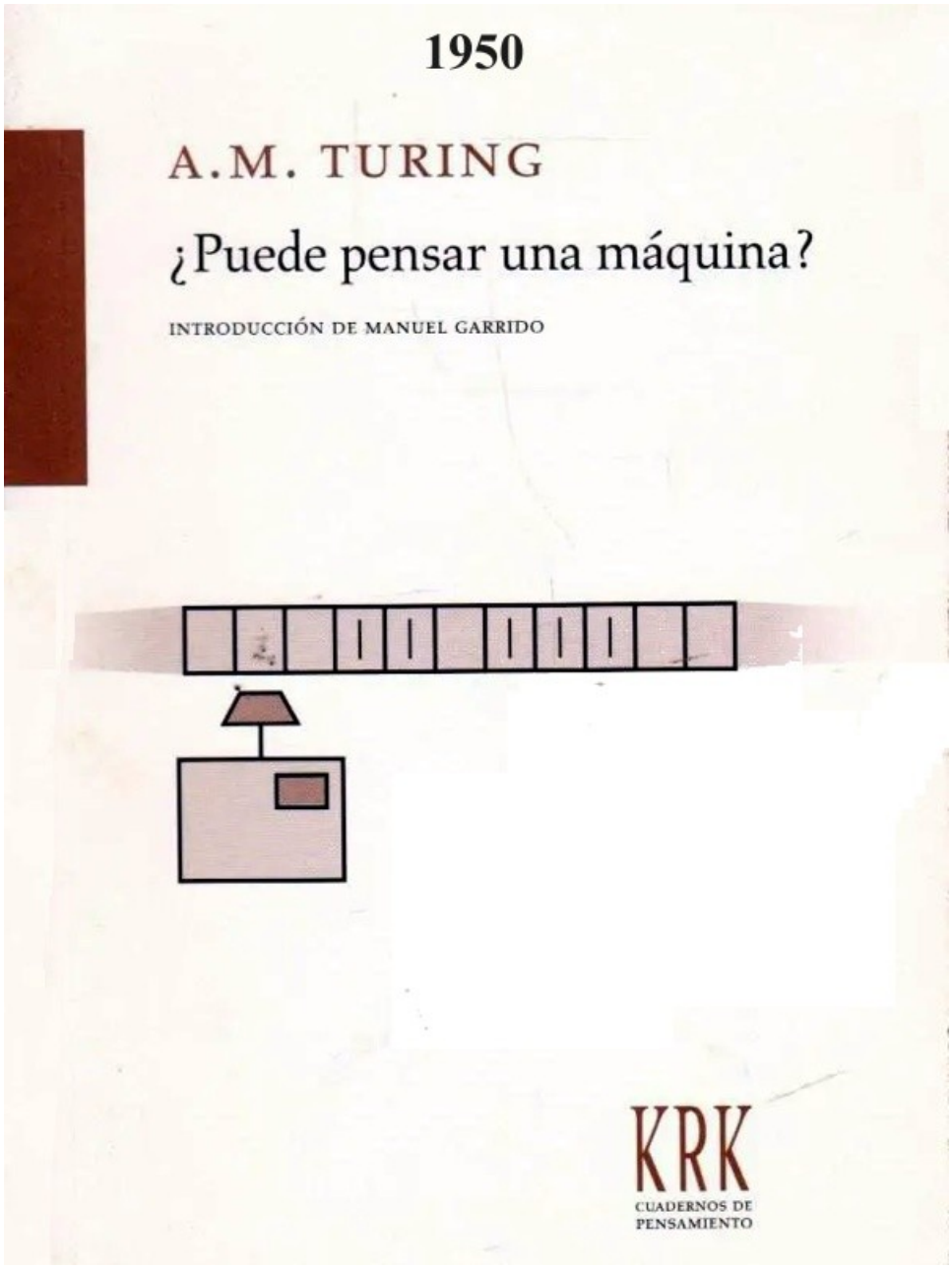
1950
A.M. TURING
¿Puede pensar una máquina?
INTRODUCCIÓN DE MANUEL GARRIDO
KRK
CUADERNOS DE PENSAMIENTO

Reseña

En 1947 Alan M. Turing pronunció una conferencia ante un auditorio compuesto en su mayor parte por miembros del National Physical Laboratory de Londres en la que intentaba responder a la vieja y controvertida pregunta ¿Puede pensar una máquina?

Lo expuesto en ese acto apareció publicado tres años más tarde en *Mind*—una importante revista de filosofía británica— y es lo que ofrecemos aquí al lector en su traducción castellana. Este texto se convirtió enseguida en uno de los escritos fundacionales de la lógica informática y la inteligencia artificial, al presentar las líneas generales por las que debería discurrir una respuesta precisa y manejable (aunque no indiscutible) a la pregunta formulada.

Se trata del famoso Test de Turing, una prueba para decidir si una máquina es inteligente (o «piensa»). Para ello Turing diseñó un juego de imitación en el que participan una máquina y seres humanos; podemos decir que una máquina piensa si un ser humano que se comunica con la máquina y con otros seres humanos no logra distinguir cuando su interlocutor es una máquina y cuando un humano.

Una «máquina de Turing» como la que participa en el juego, es un dispositivo ideal de cálculo, capaz de resolver una función computable —una función cuya solución es susceptible de ser obtenida por un procedimiento mecánico.

* * * *

Pero lo más significativo es que Turing demostró que hay una máquina peculiar —la máquina universal de Turing— en la que se puede representar cualquier máquina que sea capaz de computar una función particular. De acuerdo con esto, una máquina universal de Turing sería una especie de sistema operativo en el que se implementan diferentes programas (máquinas de Turing especiales), un poco a la manera en que nos es familiar en los ordenadores personales. La denominada «metáfora del ordenador» como modelo capaz de simular la mente humana y, por ende, el pensar, tiene aquí su fuente.

Índice

Capítulo 1
El juego de imitación

Propongo que consideremos la siguiente pregunta: « ¿Pueden pensar las máquinas?». Para empezar, definamos el significado de los términos «máquina» y «pensar», pero es una actitud peligrosa. Si hemos de llegar al significado de las palabras «máquina» y «pensar» a través de su utilización corriente, difícilmente escaparíamos a la conclusión de que hay que buscar el significado y la respuesta de la pregunta « ¿Pueden pensar las máquinas?» mediante una encuesta tipo Gallup. Pero es absurdo. En lugar de intentar tal definición, sustituiremos la pregunta por otra estrechamente relacionada con ella y que se expresa con palabras relativamente inequívocas.

El problema en su nuevo planteamiento puede exponerse en términos de un juego que denominaremos «juego de imitación». Intervienen en él tres personas: un hombre (A), una mujer (B) y un preguntador (C), indistintamente de uno u otro sexo. El preguntador se sitúa en una habitación aparte y, para él, el juego consiste en determinar quién de los otros dos es el hombre y quién la mujer. Los conoce por la referencia X e Y, y al final del juego determina si «X es A e Y es B» o si «X es B e Y es A». El preguntador puede plantear a A y a B preguntas como éstas: «Por favor X, ¿podría decirme cuán largo es su pelo?».

Supongamos que X es realmente A, entonces es A quien contesta. El objetivo de A en el juego es lograr que C efectúe una identificación errónea, por lo que su respuesta podría ser: «Mi pelo es corto,

escalonado, y los mechones más largos son de unos veinte centímetros».

Para que el preguntador no se guíe por el timbre de voz, las respuestas deben ir por escrito o, mejor aún, mecanografiadas. Lo ideal es disponer de un impresor telegráfico que comunique las dos habitaciones. Otro procedimiento consiste en que un intermediario repita pregunta y respuesta. El objeto del juego para el tercer jugador (B) es ayudar al preguntador. La mejor estrategia para la jugadora es probablemente responder la verdad, añadiendo quizás a sus respuestas cosas como ésta: « ¡Soy la mujer, no le haga caso!», pero de nada sirve, ya que el hombre puede hacer observaciones similares.

Ahora planteemos la pregunta: « ¿Qué sucede cuando una máquina sustituye a A en el juego?». ¿Se pronunciará el preguntador en este caso tan erróneamente como lo hace cuando en el juego participan un hombre y una mujer? Estas preguntas sustituyen a la original: « ¿Pueden pensar las máquinas?».

Capítulo 2
Crítica del nuevo problema

Del mismo modo que preguntamos: « ¿Cuál es la respuesta a este nuevo tipo de pregunta?», podemos preguntar: « ¿Merece la pena resolver esta nueva pregunta?». Resolvamos esta última pregunta sin plantear más objeciones para cortar una regresión infinita.

El nuevo problema presenta la ventaja de que traza una línea definida entre las aptitudes físicas e intelectuales de una persona. Ningún ingeniero o químico puede atribuirse la capacidad de producir un material que no pueda distinguirse de la piel humana. Quizá sea posible algún día, pero, aun suponiendo la viabilidad de semejante invención, nos parece que de poco serviría tratar de hacer una «máquina pensante» más humana, forrándola con esa epidermis artificial. El modo en que hemos planteado el problema refleja el obstáculo que impide al preguntador ver o tocar a los otros concursantes, oír su voz. Otras ventajas del criterio propuesto pueden resumirse en un modelo de preguntas y respuestas. Por ejemplo:

P: Por favor, escriba un soneto sobre el tema del Cuarto Puente.

R: Hágame otra pregunta; la poesía no es mi fuerte.

P: Sume 34957 con 70764.

R: (Pausa de unos 30 segundos) 105621.

P: ¿Juega al ajedrez?

R: Sí.

P: Tengo el rey en la casilla 1R y ninguna otra pieza. Usted tiene sólo el Rey en la casilla 6R y la Dama en 1D. Le toca mover. ¿Qué juega?

R: (Pausa de unos 15 segundos) La Dama a D8, mate.

El método de preguntas y respuestas parece adecuado para introducir casi todos los campos de actividad humana que queramos. No vamos a sancionar a la máquina por su incapacidad para destacar en concursos de belleza, del mismo modo que no castigamos a una persona por perder una carrera en una competición aérea. Las condiciones del juego hacen irrelevantes esas torpezas. Los «testigos» pueden alardear, si lo creen conveniente, tanto como deseen con respecto a sus encantos, su fuerza o su heroísmo, pero el preguntador no puede exigir demostraciones fehacientes.

El juego quizá provoque críticas porque la máquina tiene demasiados factores en contra. Si una persona lo intentara haciéndose pasar por la máquina, sin duda haría un papel deplorable. Quedaría rápidamente eliminada por lentitud e inexactitud aritmética. ¿No harán las máquinas algo que permita la definición de pensamiento, pero que es muy distinto a lo que hace una persona? Se trata de una objeción de peso, pero cuando menos podemos decir que, dado que es posible construir una máquina que realice satisfactoriamente el juego de imitación, la objeción no viene al caso.

Podría alegarse que la mejor estrategia en el «juego de imitación», para la máquina, es posiblemente algo distinto a la imitación de la conducta humana. Puede, pero yo no creo que esto influya demasiado. En cualquier caso, no nos proponemos aquí analizar la teoría del juego y supondremos que la mejor estrategia es tratar de dar las respuestas que una persona daría con toda naturalidad.

Capítulo 3

Las máquinas que intervienen en el juego

La cuestión que planteábamos en el apartado 1 carece de precisión si no especificamos qué entendemos por el término «máquina». Es lógico que deseemos que nuestras máquinas estén dotadas de cualquier tipo de ingeniería mecánica. Del mismo modo que aceptamos la posibilidad de que un ingeniero o un equipo de ingenieros construya una máquina que funcione, pero cuya modalidad operacional no pueden describir satisfactoriamente sus constructores porque se han servido de un método fundamentalmente experimental. Finalmente, excluiremos de la categoría de máquinas a las personas nacidas del modo habitual. Es difícil adaptar las definiciones de modo que cumplan estos tres requisitos. Se puede insistir, por ejemplo, en que el equipo de ingenieros sea de un solo sexo, lo cual no sería satisfactorio, ya que probablemente se puede crear un individuo completo a partir de una simple célula epidérmica de un hombre (pongamos por caso). Esto sería una proeza de biogenética merecedora de máxima admiración, pero no por ello la calificaríamos de «construcción de máquina pensante». Esto nos obliga a descartar el requisito de permitir cualquier tipo de técnica, y con mayor razón dado que el interés actual por las «máquinas pensantes» se ha suscitado gracias a un tipo particular de máquina, generalmente denominada «computadora electrónica» o «computadora digital». Con arreglo a

esto, sólo permitiremos que tomen parte en el juego las computadoras digitales.

A primera vista esta limitación parece muy drástica, pero intentaré demostrar que no es así. Para ello es necesario un breve resumen sobre la naturaleza y las propiedades de estas computadoras. Podría también aducirse que esta identificación de las máquinas con las computadoras digitales, al igual que nuestro criterio sobre el término «pensar», son insatisfactorias si (en contra de lo que creo) resulta que las computadoras digitales son incapaces de hacer un buen papel en el juego.

Existen ya varías computadoras operacionales, y es lógico que se diga: « ¿Por qué no realizar el experimento ahora mismo? No resultaría difícil cumplir los requisitos del juego. Se pueden utilizar varios preguntadores, compilando unas estadísticas para comprobar cuántas veces se produce la identificación correcta». La respuesta inmediata es que no se trata de plantearse si todas las computadoras digitales actuarán bien en el juego, ni de si las actuales computadoras actuarán bien, sino de si existen computadoras imaginables que actúen bien. Pero esto es sólo la respuesta inmediata, más adelante consideraremos la cuestión bajo otra perspectiva.

Capítulo 4
Computadoras digitales

Podemos explicar el concepto de computadoras digitales diciendo que son unas máquinas ideadas para realizar cualquier tipo de operación propia de un computador humano. El computador humano sigue unas reglas determinadas sin opción a desviarse de ellas bajo ningún concepto. Supongamos que esas reglas figuran en un libro que cambia cada vez que el computador acomete un nuevo trabajo. Dispone también de una cantidad ilimitada de papel para efectuar cálculos y hace las multiplicaciones y sumas pertinentes con una «máquina de bolsillo», pero esto no tiene importancia.
Si utilizamos como definición la anterior explicación, corremos el riesgo de caer en una argumentación circular. Para evitarlo, esbozaremos los medios con los que se logra el efecto deseado. Suele considerarse que una computadora digital consta de tres partes:

1. *Almacenamiento*
2. *Unidad procesadora*
3. *Control*

El almacenamiento es el acopio de información y corresponde al papel sobre el que se efectúa la computación humana, ya sea el papel en que la persona realiza los cálculos o aquél en el cual está impreso el libro de reglas. Del mismo modo que el computador humano efectúa sus cálculos con su cabeza, parte del almacenamiento corresponde a la memoria de la máquina.

La unidad procesadora es el sector que realiza las distintas operaciones de cálculo. La naturaleza de estas operaciones varía de una máquina a otra. Generalmente pueden efectuar operaciones bastante largas, tales como «Multiplicar 3540675445 por 7076345687», pero en algunas máquinas sólo pueden llevarse a cabo operaciones muy simples, tales como «Escribe 0».

Hemos mencionado que el «libro de reglas», de que se vale el computador, se sustituye en la máquina por una parte del almacenamiento. Esta se denomina «tabla de instrucciones». Corresponde al control comprobar que las instrucciones se sigan correctamente y en su debido orden. El control está construido de tal manera que es infalible.

La información almacenada suele estar dividida en paquetes de tamaño relativamente modesto. En una máquina concreta, por ejemplo, el paquete puede constar de diez dígitos decimales. Se asignan números a las partes del almacenamiento en que se guardan los diversos paquetes de información, con arreglo a una modalidad sistemática. Un ejemplo de instrucción corriente podría ser: «Suma la cifra almacenada en la posición 6809 a la situada en la 4302 y devuelve el resultado de la última posición de almacenamiento». Ni que decir tiene que la operación no se desarrolla en la máquina expresada de este modo, sino que se lleva a cabo siguiendo una codificación como 6809430217. La cifra 17 indica cuál de las posibles operaciones hay que efectuar con las dos cifras. En cuyo caso la operación es la anteriormente descrita: «Suma la cifra...». Se advertirá que la instrucción consta de diez

dígitos y, por lo tanto, constituye exactamente un paquete informativo. El control suele captar las instrucciones a seguir en el orden de posición en que están almacenadas, aunque a veces pueda surgir una instrucción como ésta: «Sigue ahora la instrucción almacenada en la posición 5606 y continúa», o bien: «Si la posición 4505 contiene 0, sigue la instrucción almacenada en 6707; en caso contrario continúa».

Las instrucciones de este tipo son muy importantes porque permiten la repetición de una secuencia de operaciones una y otra vez hasta que se cumple un determinado requisito, pero, al hacerlo, la máquina sigue en cada repetición, no nuevas instrucciones, sino las mismas indefinidamente. Recurramos a una analogía casera: supongamos que mamá desea que Tommy pase por el zapatero cada mañana camino del colegio para ver si han arreglado sus zapatos; puede decírselo cada mañana, o puede dejar una nota permanente en el vestíbulo para que el niño la vea al salir y recuerde que tiene que pasar por el zapatero, y luego, al volver, si trae los zapatos, rompa la nota. El lector debe aceptar como un hecho la construcción de computadoras digitales que, efectivamente, se han construido con arreglo a los principios expuestos y que realmente mimetizan con gran fidelidad los actos de un computador humano.

El libro de reglas que, según hemos señalado, utiliza el computador humano es, naturalmente, una ficción convencional. Los computadores humanos recuerdan en realidad lo que tienen que hacer. Si queremos hacer una máquina que mimetice el comportamiento de un computador humano en operaciones

complicadas, hay que preguntarle a éste cómo lo hace y luego transferir la respuesta en forma de tabla de instrucciones. La elaboración de tablas de instrucciones suele denominarse «programación». La «programación de una máquina para que efectúe la operación A» significa insertar en la máquina la tabla de instrucción adecuada para que lleve a cabo A.

Una variante interesante de la idea de computadora digital es la «computadora digital con un elemento aleatorio». Estas máquinas disponen de instrucciones en las que interviene un dado o un proceso electrónico equivalente; una instrucción de este tipo puede ser, por ejemplo: «Arroja el dado y almacena la cifra resultante en 1000». A veces se las denomina máquinas de libre voluntad (aunque personalmente yo no utilice esta expresión). Normalmente no se puede determinar por simple observación de la máquina si ésta posee un elemento aleatorio, ya que se logra un efecto similar con dispositivos cuya elección depende de los dígitos de los decimales de π.

La mayoría de las computadoras digitales poseen un almacenamiento finito, aunque no existe dificultad teórica en la concepción de una computadora de almacenamiento ilimitado. Naturalmente, sólo podría utilizarse una parte finita de cada fase. De igual modo se habría podido construir una cantidad finita, pero cabe imaginar que sucesivamente fueran añadiéndose otras. Estas computadoras presentan especial interés teórico y las denominaremos computadoras de capacidad infinita.

El concepto de computadora digital es antiguo. Charles Babbage, profesor de matemáticas en la Universidad de Cambridge entre 1828 y 1839 concibió una a la que denominó Máquina Analítica, pero no la terminó. Aunque Babbage expuso los principios fundamentales, la máquina no representaba en aquella época gran interés. Su rapidez habría sido mucho mayor que la de un computador humano, pero unas 100 veces inferior a la de la máquina de Manchester, que a su vez es una de las máquinas modernas más lentas. El almacenamiento era puramente mecánico y se efectuaba por medio de ruedas y tarjetas.

El hecho de que la Máquina Analítica de Babbage estuviera concebida de forma totalmente mecánica nos ayudará a despejar cualquier superstición. Muchas veces se atribuye importancia al hecho de que las computadoras digitales modernas son eléctricas, igual que el sistema nervioso. Como la máquina de Babbage no era eléctrica, y como todas las computadoras digitales son en cierto modo equivalentes a ella, el empleo de la electricidad no es teóricamente relevante.

Siempre que se trata de señalización rápida interviene, claro, la electricidad. Por lo tanto, no es de extrañar que ésta se halle relacionada con ambos conceptos. En el sistema nervioso los fenómenos químicos son, cuando menos, tan importantes como los eléctricos. En ciertas computadoras el sistema de almacenamiento es fundamentalmente acústico. Por lo tanto, el empleo de la electricidad como propiedad no deja de ser una similitud muy

superficial. Para establecer similitudes reales debemos más bien buscar analogías en el funcionamiento matemático.

Capítulo 5
Universalidad de las computadoras digitales

Podemos situar las computadoras digitales que hemos tratado en el apartado anterior dentro de la categoría de «máquinas de estado discreto». Estas son máquinas que pasan mediante saltos o clics súbitos de un estado bastante definido a otro. Se trata de estados lo bastante distintos para que no se dé la posibilidad de confusión entre ellos. Hablando en puridad no existen tales máquinas. En realidad, todo se mueve continuamente, pero podemos considerar positivamente muchos tipos de máquinas como de estado discreto. Por ejemplo, al referirnos a los interruptores de un sistema de iluminación, es una ficción convencional decir que cada uno de ellos debe hallarse totalmente conectado o desconectado. Pueden hallarse en posiciones intermedias, pero en la mayoría de los casos podemos descartarlas. Como ejemplo de máquina de estado discreto consideremos una rueda que recorra 120° por segundo, pero que se detiene al accionar una palanca externa; ésta, además, en determinada posición, enciende una luz. Podríamos definir esta máquina de forma abstracta del siguiente modo: El estado interno de la máquina (descrito por la posición de la rueda) puede ser q_1, q_2 o q_3. Hay una señal de entrada i_0 o i_1 (posición de la palanca). El estado interno en cualquier momento está determinado por el último estado, y la señal de entrada lo estará con arreglo a la tabla:

		Último Estado		
		q_1	q_2	q_3
Entrada	i_0	q_2	q_3	q_1
	i_1	q_1	q_2	q_3

Las señales de salida, única indicación visible externa del estado interno (la luz), nos las da la tabla

Estado	q_1	q_2	q_3
Salida	o_0	o_0	o_1

Es un ejemplo clásico de máquina de estado discreto. Este tipo de máquinas se describen por medio de las tablas indicadas, a condición de que posean únicamente un número finito de estados posibles.

Podría parecer que, dado el estado inicial de la máquina y la señal de entrada, siempre fuera posible predecir los estados futuros, pero es una reminiscencia de la perspectiva de Laplace, según la cual, a partir del estado completo del universo en un momento del tiempo, definido por las posiciones y velocidades de todas sus partículas, se pueden predecir los estados futuros. Sin embargo, la predicción que estamos considerando es más próxima a la practicabilidad que la considerada por Laplace. El sistema del «universo como un todo» es de tal naturaleza que errores bastante pequeños en las condiciones

iniciales pueden ejercer un efecto considerable en un momento futuro. El desplazamiento de un solo electrón en una billonésima de centímetro en un momento determinado puede ser la causa de que una persona muera aplastada por una avalancha un año más tarde o se libre de la catástrofe. Es una propiedad esencial de los sistemas mecánicos, que hemos denominado «máquinas de estado discreto», el que semejante fenómeno no se produzca. Incluso si consideramos las actuales máquinas fisicas en lugar de las máquinas idealizadas, el conocimiento razonablemente exacto de su estado en determinado momento nos procura un conocimiento razonablemente exacto de cualquier serie de pasos ulteriores.

Como hemos dicho, las computadoras digitales pertenecen al grupo de máquinas de estado discreto. Pero el número de estados que pueden adoptar este tipo de máquinas suele ser enormemente elevado. Por ejemplo, para la máquina que actualmente funciona en Manchester, la cifra aproximada sería de 2^{165000}, es decir de 10^{50000}aproximadamente. Compárese esto con el citado ejemplo de la rueda que tenía tres estados. Se comprende sin dificultad por qué es tan elevado el número de estados. La computadora posee un almacenamiento correspondiente al papel que utiliza un computador humano. En este almacenamiento puede escribirse cualquiera de las combinaciones de símbolos que figurasen en el papel. Para simplificar, supongamos que sólo utilizamos como símbolos los dígitos del 0 al 9. No tomaremos en cuenta las variaciones de los signos manuscritos. Supongamos que la computadora dispone de 100 hojas de papel de 50 líneas cada una,

con espacio para 30 dígitos. El número de estados será $10^{10\times50\times30}$, es decir, 10^{150000}. Esto equivale aproximadamente al número de estados de tres máquinas de Manchester juntas. El logaritmo con base dos del número de estados es en realidad lo que se denomina «capacidad de almacenamiento» de la máquina. Por lo tanto, la máquina de Manchester posee una capacidad de almacenamiento aproximada de 165000, y la máquina con rueda del ejemplo mencionado, de aproximadamente 1,6. Si juntamos dos máquinas, habrá que sumar sus capacidades para saber la capacidad de la máquina resultante. Esto nos permite afirmar que «la máquina de Manchester contiene 64 pistas magnéticas, cada una de ellas con capacidad para 2560, ocho tubos electrónicos con capacidad de 1280. El almacenamiento diverso equivale aproximadamente a 300, lo que da un total de 174380».

Disponiendo de la tabla correspondiente a una máquina de estado discreto se puede predecir lo que hará, y nada nos impide efectuar este cálculo con una computadora digital. A condición de que lo efectúe con suficiente rapidez, la computadora digital puede mimetizar el comportamiento de cualquier máquina de estado discreto. Entonces, se podría jugar con esa máquina (en el papel B) al juego de imitación y con la computadora digital mimetizante (en el papel de A), y el interrogador no sabría diferenciarlas. Naturalmente, la computadora digital debe poseer una capacidad de almacenamiento adecuada y funcionar a suficiente velocidad. Además, habrá que programarla expresamente para cada nueva máquina que se desee imitar.

Esta propiedad esencial de las computadoras digitales, por la que pueden imitar a cualquier máquina de estado discreto, se define diciendo que son máquinas *universales*. La existencia de máquinas con esta propiedad encierra la importante consecuencia de que, consideraciones de rapidez aparte, no hay necesidad de diseñar diversas máquinas nuevas para que realicen los correspondientes nuevos procesos de computación. Todos pueden efectuarse con una sola computadora digital, convenientemente programada en cada caso. En consecuencia, como veremos, todas las computadoras digitales de este tipo son equivalentes en un sentido.

Ahora consideraremos la cuestión mencionada al final del apartado 3. Habíamos sugerido sustituir la pregunta « ¿Pueden pensar las máquinas?» por la de « ¿Existen computadoras digitales imaginables que jueguen bien al juego de imitación?». Si se desea, puede generalizarse más superficialmente esta pregunta: « ¿Hay máquinas de estado discreto que hagan un buen juego?». Pero, dada la propiedad universal, vemos que ambas preguntas equivalen a: «Supongamos una determinada computadora digital C. ¿Es cierto que, modificando esta computadora para que tenga un almacenamiento adecuado y dotándola de un programa apropiado, podemos conseguir que C desempeñe eficazmente el papel de A en el juego de imitación y el papel de B lo haga un hombre?».

Capítulo 6

Opiniones contrapuestas sobre la cuestión principal

Contenido:

Consideremos ahora que hemos despejado el terreno y podemos ya pasar al debate de la pregunta « ¿Pueden pensar las máquinas?» y de su variante, expuesta al final del apartado anterior. No podemos descartar totalmente la forma original del problema, ya que habrá diversidad de opiniones con respecto a la pertinencia de la sustitución y no podemos por menos que atender lo que se diga sobre el asunto.

Simplificaré las cosas para el lector si, en primer lugar, explico mi propia opinión sobre el tema. Consideremos primero la forma más exacta de la pregunta. Personalmente creo que, dentro de unos cincuenta años, se podrá perfectamente programar computadoras con una capacidad de almacenamiento aproximada de 10^9 para

hacerlas jugar tan bien al juego de imitación que un preguntador corriente no dispondrá de más del 70 por ciento de las posibilidades para efectuar una identificación correcta a los cinco minutos de plantear las preguntas. Me parece que la pregunta original, « ¿Pueden pensar las máquinas?», no merece discusión por carecer de sentido. No obstante, creo que, a finales del siglo, el sentido de las palabras y la opinión profesional habrán cambiado tanto que podrá hablarse de máquinas pensantes sin levantar controversias. Creo además que de nada sirve ocultar las ideas. La opinión tan generalizada de que los científicos proceden siempre de un hecho bien demostrado a otro hecho bien demostrado, y nunca se dejan influir por una conjetura no probada, es bastante errónea. A condición de que quede bien claro qué son hechos probados y qué son conjeturas, no existe ningún peligro. Las conjeturas son de suma importancia, porque sugieren posibles vías de investigación.
Ahora consideraré opiniones contrarias a la mía:

§. La objeción teológica

El pensamiento es una función del alma inmortal del hombre. Dios ha dado un alma inmortal a todos los hombres y mujeres, pero no a ningún animal ni máquina. Por lo tanto, ni los animales ni las máquinas pueden pensar.
Personalmente son ideas que rechazo totalmente, pero intentaré refutarlas en términos teológicos. La argumentación resultaría más convincente si se clasificara a los animales con el hombre, ya que existe mucha más diferencia, para mí, entre lo genuinamente

animado y lo inanimado que entre el hombre y los animales. El carácter arbitrario de la opinión ortodoxa se evidencia aún más si tenemos en cuenta la opinión de los creyentes de otras religiones. ¿Cómo ve el cristianismo el dogma musulmán según el cual la mujer no tiene alma? Pero dejemos esto y volvamos a la cuestión principal. Creo que el citado argumento implica una grave restricción de la omnipotencia del Todopoderoso. Se admite así que hay cosas de las que Él es incapaz, como es hacer que uno sea igual a dos, pero ¿dudaremos de su libertad para insuflar alma a un elefante, si a bien lo tiene? Cabe esperar que únicamente ejerciese tal poder en conjunción con una mutación que dotase al elefante de un cerebro mejorado que respondiera a las necesidades de esa alma. Podemos argüir exactamente lo mismo en el caso de las máquinas. Puede parecer distinto por ser más difícil de «tragar», pero esto únicamente significa que pensamos que es menos verosímil que Él considere adecuadas las circunstancias para dotarlas de alma. Las circunstancias en cuestión se discuten en el resto de este trabajo. Al intentar construir este tipo de máquinas no estamos usurpando irreverentemente Su poder de crear almas, igual que no lo hacemos al procrear niños; en realidad, en ambos casos somos instrumentos de Su voluntad al procurar moradas para las almas que Él crea.

Pero todo esto es mera especulación. No me impresionan mucho los argumentos teológicos, aunque se utilicen como apoyo. A lo largo de la historia se ha comprobado cuánto dejan que desear. En tiempos de Galileo se argumentaba que las Sagradas Escrituras decían: «Y el

sol se detuvo… y no fue hacia el ocaso durante casi un día» (Josué x.13) y que: «Él creó los fundamentos de la Tierra para que no se moviera» (Salmo cv. 5) como refutación convincente de la teoría copernicana. Con los conocimientos actuales estos argumentos resultan fútiles, pero en una época de escasos conocimientos científicos causaban muy distinta impresión.

§. La objeción del «avestruz»

«Las consecuencias de que las máquinas piensen serían horribles. Creamos y esperemos que no sea posible».

Este argumento rara vez se expone de forma tan abierta, pero afecta a la mayoría de quienes reflexionamos sobre ello. Nos gusta creer que el hombre es en algún modo superior al resto de la creación, y tanto mejor si podemos demostrar que es necesariamente superior, pues entonces no existe peligro de que pierda su posición dominante. La popularidad del argumento teológico está claramente vinculada a esta idea y cuenta con muchos adeptos entre los intelectuales, pues éstos aprecian más que otras personas el poder del pensamiento y se muestran más inclinados a basar su convencimiento de la superioridad del hombre en este poder.

No creo que este argumento sea lo bastante fundado para molestarme en refutarlo. Tal vez sea mejor consolarse, buscándolo quizás en la transmigración de las almas.

§. La objeción matemática

Pueden citarse toda una serie de resultados de la lógica matemática para demostrar que hay limitaciones en el poder de las máquinas de estado discreto. El más conocido es el denominado teorema de Gödel, que demuestra que en cualquier sistema lógico lo bastante potente pueden formularse afirmaciones que no pueden demostrarse ni refutarse dentro del sistema, salvo en caso de que posiblemente tal sistema sea incoherente. En ciertos aspectos similares hay otros resultados expuestos por Church, Kleene, Rosser y Turing. La tesis de este último autor es la que merece mayor consideración en este caso, por referirse específicamente a las máquinas, mientras que las de los otros sólo son utilizables en tanto que argumentos relativamente indirectos: si, por ejemplo, recurrimos al teorema de Gödel, hace falta a la vez disponer de medios para describir los sistemas lógicos en términos de máquinas y las máquinas en términos de sistemas lógicos. El resultado en cuestión se refiere a un tipo de máquina que es fundamentalmente una computadora digital con capacidad infinita, y postula que hay ciertas cosas que esa máquina no puede efectuar. Si se la equipa para dar respuesta a preguntas como en el juego de imitación, habrá preguntas que contestará mal o no podrá contestar por mucho tiempo que se le conceda. Naturalmente, puede haber muchas preguntas de esta clase, y preguntas que no pueda contestar satisfactoriamente una máquina las contestará adecuadamente otra. Desde luego estamos por ahora en la suposición de que estas preguntas son de tal índole que la respuesta es «Sí» o «No», y no preguntas del tipo « ¿Qué opinas sobre

Picasso?». Las preguntas que sabemos que la máquina no contesta son de esta clase: «Supongamos una máquina con las siguientes características... ¿contestará esta máquina "Sí" a cualquier pregunta?». Los puntos suspensivos se sustituyen por la descripción de una máquina modelo estándar, que podría ser como la que se cita en el apartado 5. Si la máquina descrita guarda cierta relación comparativamente simple con la máquina a que se está interrogando, puede demostrarse que la respuesta es incorrecta o no se va a producir. Este es el resultado matemático: se arguye que demuestra una incapacidad por parte de las máquinas a la que no está expuesto el intelecto humano.

La respuesta taxativa a este razonamiento es que, aunque está demostrado que existen limitaciones en la capacidad de cualquier máquina, sólo se ha afirmado, sin prueba alguna, que tales limitaciones no son aplicables al intelecto humano. Sin embargo, yo no creo que esta posibilidad pueda rechazarse tan alegremente. Cuando se plantea a una de estas máquinas la pregunta crítica adecuada y nos da una respuesta concreta, sabemos que la respuesta es incorrecta y esto nos da cierta sensación de superioridad. ¿Es una sensación ilusoria? Sin duda es lo bastante legítima, pero yo no creo que haya que atribuirle demasiada importancia. También nosotros en muchas ocasiones respondemos erróneamente a preguntas, lo cual no justifica esa enorme sensación de halago al ver que las máquinas fallan. Además, sólo podemos sentir en este caso nuestra superioridad en relación con la máquina concreta, objeto de nuestra frágil victoria. No es un triunfo

simultáneo frente a *todas* las máquinas. En resumen, habrá hombres más listos que cualquier máquina, pero también otras máquinas más listas, y así sucesivamente.

Los partidarios del argumento matemático aceptarán en su mayoría —creo yo— que el juego de imitación es una buena base para la discusión. A los partidarios de las dos primeras objeciones seguramente no les interesará ningún razonamiento.

§. El argumento de la conciencia

Este argumento está perfectamente expresado en un discurso conmemorativo del profesor Jefferson, en 1949, del que cito: «Hasta que una máquina sea capaz de escribir un soneto o de componer un concierto, porque tenga la facultad de reflexionar y sea capaz de sentir, y no por la combinación aleatoria de símbolos, no podremos admitir que esa máquina sea igual al cerebro, en el sentido de que no sólo los escriba, sino que sepa que los ha escrito. Ningún mecanismo (y no hablo de una señal artificial, invención simplona) puede sentir placer por sus logros, pena cuando se funden sus válvulas, regocijo por los halagos, depresión por sus errores, atracción sexual, enfado o decepción cuando no consigue lo que quiere».

Este argumento parece ser la negación de la validez de nuestro test. Según la modalidad más extremada de este tipo de planteamiento, la única manera de asegurarse de que una máquina piensa es ser la máquina y sentir el propio pensamiento. Sólo entonces pueden exponerse tales sentimientos a todo el mundo, pero tampoco está

justificado que a nadie le importen. Según este planteamiento, también la única manera de saber que una persona piensa es ser esa persona concreta. De hecho, es un punto de vista solipsista. Puede que sea el punto de vista más lógico, pero dificulta la comunicación de ideas. A puede sentirse inclinado a creer «A piensa pero B no», mientras que B creerá que «B piensa pero A no». En lugar de discutir indefinidamente este punto, mejor es adscribirse al cortés convencionalismo de que todos piensan.

Estoy convencido de que el profesor Jefferson no desea adoptar el punto de vista extremo y solipsista. Probablemente se halle dispuesto a aceptar como prueba el juego de imitación. El test (omitiendo el jugador B) suele usarse en la práctica bajo la denominación de *examen oral* para descubrir si el candidato entiende de verdad algo o lo «ha aprendido como un papagayo». Escuchemos un extracto de uno de esos *exámenes orales*:

> ***Examinador:*** *En el primer verso de su soneto, que dice « ¿Te compararía con un día de verano?», ¿no sería igual, o mejor, «un día de primavera»?*
>
> ***Examinado:*** *No rimaría.*
>
> ***Examinador:*** *¿Y «un día de invierno»? Rima perfectamente.*
>
> ***Examinado:*** *Sí, pero a nadie le gusta que le comparen con un día de invierno.*
>
> ***Examinador:*** *¿Diría usted que Mr. Pickwick le recuerda la Navidad?*
>
> ***Examinado:*** *En cierto modo.*

Examinador: *Pues Navidad es un día de invierno, y no creo que a Mr. Pickwick le molestara la comparación.*

Examinado: *Creo que bromea usted. Por día de invierno se entiende un día de invierno genuino y no uno especial como el de Navidad.*

* * * *

Y así sucesivamente. ¿Qué diría el profesor Jefferson si la máquina escritora de sonetos fuera capaz de contestar así en el examen oral? No sé si la consideraría accionada por «una simple señal artificial» al dar tales respuestas, pero, si las respuestas fueran tan adecuadas y coherentes como en el párrafo anterior, no creo que las calificara de «invención simplona». Yo creo que con esta expresión se intenta definir dispositivos tales como la inclusión en la máquina de un disco de alguien que lee un soneto, dotado del correspondiente relé que lo conecte de vez en cuando.

En resumen, creo que a la mayoría de los partidarios del argumento de la conciencia se les podría convencer de que lo abandonaran en lugar de forzarles a la actitud solipsista. Entonces, probablemente se inclinasen a aceptar la prueba.

No quisiera dar la impresión de que creo que no existe misterio en lo que se refiere a la conciencia. Existe, por ejemplo, algo así como una paradoja en relación con su localización. Pero no creo que haya que solucionar necesariamente ese misterio para responder a la cuestión que nos ocupa en este trabajo.

§. Argumentos de incapacidades diversas

Estos argumentos responden al esquema: «Te aseguro que pueden hacerse máquinas que realicen todo lo que has dicho, pero es imposible construir una máquina que haga X», y se citan al respecto diversas X. A continuación expongo una selección:

«Ser amable, ingeniosa, hermosa, amistosa», «poseer iniciativa, tener sentido del humor, distinguir entre lo bueno y lo malo, cometer faltas», «enamorarse, apreciar las fresas y los helados», «enamorar a alguien, aprender por la experiencia», «utilizar adecuadamente las palabras, ser objeto de su propio pensamiento», «tener un comportamiento tan versátil como una persona, hacer algo auténticamente nuevo».

Generalmente estas afirmaciones no se apoyan en razonamientos y, personalmente, creo que en esencia se basan en el principio de la inducción científica. Una persona ve miles de máquinas durante su vida y, por lo que ve de ellas, extrae una serie de conclusiones generales. Son feas y cada una de ellas está ideada para una tarea concreta; cuando se desea que ejecuten varias funciones, son inservibles, su variedad de comportamiento es muy limitada, etc., etc. En consecuencia, concluye que ésas son las características de las máquinas en general. Muchas de estas limitaciones se asocian a la escasa capacidad de almacenamiento de la mayoría de las máquinas (supongo que, en el concepto de capacidad de almacenamiento, se incluyen en cierto modo a las máquinas distintas a las de estado discreto. No importa la definición exacta,

ya que no aspiramos a una exactitud matemática en esta discusión). Hace unos años, cuando aún se hablaba poco de computadoras digitales, era de esperar que su mención suscitara incredibilidad cuando se hablaba de sus propiedades sin explicar su construcción. Supongo que era también debido a la aplicación del principio de inducción científica. Naturalmente esta clase de aplicación del principio suele ser inconsciente. Cuando un niño que ha sufrido una quemadura teme al fuego y demuestra que lo teme evitándolo, decimos que está aplicando la inducción científica. (Naturalmente, puedo también describir su comportamiento de muchas otras maneras). Los trabajos y las costumbres humanos no parecen constituir un material muy adecuado para la aplicación de la inducción científica. Habría que investigar una gran magnitud espacio-temporal para obtener resultados fiables, pues, si no, creeremos (como la inmensa mayoría de los niños ingleses) que todo el mundo habla inglés y que es una tontería aprender francés.

Sin embargo, conviene hacer algunas observaciones respecto de las múltiples incapacidades que hemos citado. La incapacidad para apreciar las fresas y los helados le habrá parecido al lector una futilidad. Puede que se construya una máquina que aprecie esos manjares, pero sería una imbecilidad intentarlo. Lo importante respecto de esta incapacidad es que está destinada a aumentar el número de incapacidades, por ejemplo, el mismo tipo de dificultad de comunicación amistosa que se produce entre el hombre y la máquina también se da entre un hombre blanco y otro hombre blanco, o entre un hombre negro y otro hombre negro.

Afirmar que las «máquinas no cometen errores» parece curioso. Se siente uno inclinado a replicar: « ¿Y son por eso peores?», pero adoptemos una actitud más simpática y tratemos de comprender qué es lo que significa. Creo que esta crítica puede explicarse en términos del juego de imitación. Se afirma que al preguntador le basta, para distinguir una máquina del hombre, plantear una serie de problemas aritméticos. La máquina queda desenmascarada por su tremenda exactitud. Así de sencillo, pero la máquina (programada para jugar el juego) no tratará de dar las respuestas *correctas* a los problemas aritméticos e introducirá deliberadamente errores de modo calculado para confundir al preguntador. Una avería mecánica se percibirá probablemente al darse una decisión inadecuada respecto del tipo de error aritmético a efectuar. Incluso esta interpretación crítica no es lo bastante simpática, pero no disponemos de espacio para extendernos más. A mí me parece que la crítica se fundamenta en una confusión de dos tipos de error. Podemos denominarlos «errores de funcionamiento» y «errores de conclusión». Los errores de funcionamiento los causa un efecto mecánico o eléctrico que obliga a la máquina a comportarse de modo distinto a como está diseñada. En las discusiones filosóficas se ignora la posibilidad de tales errores y se habla de «máquinas abstractas». Estas máquinas abstractas son ficciones matemáticas más que objetos físicos. Son, por definición, incapaces de errores de funcionamiento. En este sentido podemos afirmar con certeza que «las máquinas no cometen errores». Los errores de conclusión sólo pueden producirse cuando se atribuye un significado a las señales

de salida de la máquina. La máquina puede, por ejemplo, imprimir ecuaciones matemáticas, o frases en inglés. Cuando escribe una oración incorrecta, decimos que ha cometido un error de conclusión. Evidentemente no existe motivo para decir que una máquina no puede cometer este tipo de error. Puede que se limite a escribir sin parar «0 = 1». Adoptando un ejemplo menos peyorativo, digamos que, al estar dotada de un método para extraer conclusiones por inducción científica, es presumible que semejante método conduzca a veces a resultados erróneos.

A la afirmación de que una máquina no puede ser objeto de su propio pensamiento sólo puede contestarse si se demuestra que la máquina posee algún pensamiento referido a *algún* tema. No obstante, «el tema de las operaciones de una máquina» parece significar algo, al menos para quienes trabajan con ella. Si, por ejemplo, la máquina trata de hallar la solución a la ecuación $x^2 - 40x - 11 = 0$, uno no puede resistir la tentación de calificar esta ecuación de objeto parcial del tema de la máquina en ese momento. En este aspecto no cabe duda de que una máquina es su propio objeto, ya que se la puede utilizar para que contribuya a la confección de su propio programa, o para predecir el efecto de alteraciones en su propia estructura. Observando los resultados de su propio comportamiento, es capaz de modificar sus programas para efectuar determinada tarea con mayor eficacia. Son posibilidades de un futuro no muy lejano, no sueños utópicos.

La crítica de que una máquina no puede tener versatilidad de comportamiento es sólo una manera de decir que no puede tener

una gran capacidad de almacenamiento. Hasta hace relativamente poco tiempo una simple capacidad de mil dígitos era algo extraordinario.

Las críticas que estamos considerando suelen ser variantes enmascaradas del argumento de la conciencia. Generalmente, si uno sostiene que una máquina *puede* hacer una de esas cosas y describe la clase de método del que puede servirse, no se logra impresionar a los detractores, pues piensan que el método (sea el que fuere, por ser mecánico necesariamente) es algo vil. Cotéjese el paréntesis del párrafo de Jefferson citado anteriormente.

§. Objeción de lady Lovelace

La información más pormenorizada sobre la máquina analítica de Babbage figura en un informe de lady Lovelace. En él se afirma: «La Máquina Analítica no pretende *crear* nada. Puede realizar *lo que nosotros sepamos mandarle*» (en cursiva en el informe original). Es Hartree quien cita este párrafo, y añade: «Esto no implica que sea imposible construir equipo electrónico que "piense por sí solo", o en el que, en términos biológicos, no se pueda implantar un reflejo condicionado que sirva de base al "aprendizaje". Si es o no posible en principio, es una cuestión apasionante y estimulante, esbozada en algunos de los últimos avances tecnológicos. Pero no parecía que las máquinas construidas en aquella época tuvieran tal propiedad».

Coincido totalmente con Hartree al respecto. Adviértase que él no afirma que la máquina en cuestión no posea la propiedad, sino que a lady Lovelace no le constaba que la tuviera. Es muy posible que

las máquinas en cuestión tuvieran en cierto modo esa propiedad. Supongamos que una máquina de estado discreto tiene esa propiedad. La Máquina Analítica era una computadora digital universal, de forma que, si su capacidad de almacenamiento y su velocidad eran adecuados, con un programa idóneo se la podría inducir a mimetizar la propia máquina. Probablemente este razonamiento no se le ocurrió a la condesa ni al propio Babbage. En cualquier caso, ellos no tenían por qué reivindicar todo lo reivindicable.

Volveremos a hablar del tema en el apartado de máquinas que aprenden.

Una variante a la objeción de lady Lovelace afirma que las máquinas «nunca hacen nada nuevo». Podemos parangonar tal afirmación al refrán: «No hay nada nuevo bajo el sol». ¿Quién puede tener el firme convencimiento de que el «trabajo original» que se acaba de realizar no es sino el desarrollo de la simiente que ha dejado en él el aprendizaje, o la consecuencia de atenerse a consabidos principios generales? Otra variante mejor de esta objeción es la de que la máquina nunca «puede sorprendernos». Es un desplante más directo, por lo que respondemos directamente. Las máquinas me sorprenden muy a menudo. Fundamentalmente porque no calculo lo suficiente para figurarme lo que van a hacer, o, más bien, porque, aunque calculo, lo hago de forma precipitada, descuidada y corriendo riesgos, y me digo: «Supongo que el voltaje es aquí el mismo que allí; bueno, supongamos que es el mismo». Naturalmente, muchas veces me equivoco, el resultado me

sorprende, aunque, una vez finalizado el experimento, me olvido de mis falsas suposiciones. Con esta confesión me expongo a sermones sobre mis malas costumbres, pero no empaño mi sinceridad al dar fe de las sorpresas que experimento.

No pretendo con esta réplica silenciar la crítica. Probablemente puede deducirse que tales sorpresas se deben a algún acto creativo mental por mi parte, y nada dicen a favor de la máquina. Esto nos obliga a volver al argumento de la conciencia, muy lejos de la idea de sorpresa. Es un tipo de argumentación muy similar, pero quizá valga la pena señalar que la apreciación de algo como sorprendente requiere igual «acto mental creativo», independientemente de que la sorpresa la cause una persona, un libro, una máquina o lo que sea.

La opinión de que las máquinas no pueden producir sorpresa se basa, creo yo, en el sofisma en el que suelen incurrir particularmente filósofos y matemáticos: la asunción de que, cuando a la mente se le presenta un hecho, todas las consecuencias del mismo la invaden con él simultáneamente. Es una asunción muy útil en muchas circunstancias, pero se olvida con harta facilidad de que es falsa. Una consecuencia natural de asumirla como cierta es que se da por sentado que no hay mérito en la simple elucidación de consecuencias a partir de datos y principios generales.

§. Argumento de la continuidad del sistema nervioso

Desde luego el sistema nervioso no es una máquina de estado discreto. Un pequeño error de información sobre la magnitud de un

impulso nervioso aferente en una neurona puede modificar considerablemente la magnitud del impulso de salida. Puede argüirse que, precisamente por eso, no cabe posibilidad de mimetizar el comportamiento del sistema nervioso mediante un sistema de estado discreto. Cierto es que una máquina de estado discreto es distinta a una máquina continua, pero, si nos ceñimos a las condiciones del juego de imitación, el preguntador no gana nada con esa diferencia. Podemos aclarar la situación si consideramos cualquier otra máquina continua más sencilla. Un analizador diferencial, por ejemplo. (Un analizador diferencial es un tipo de máquina de estado no discreto que se emplea para cierta clase de cálculos). Algunos dan la respuesta impresa, por lo que son adecuados para intervenir en el juego. Una computadora digital no puede predecir exactamente las respuestas que da a un problema el analizador diferencial, pero sí puede dar la respuesta correcta. Por ejemplo, si se pregunta el valor de π (3,1416 aproximadamente), es razonable elegir al azar entre los valores 3'12, 3'13, 3'14, 3'15, 3'16 con las probabilidades de 0'05, 0'15, 0'55, 0'19, 0'06 (pongamos por caso). En tales circunstancias resultará muy difícil para el preguntador distinguir al analizador diferencial de la computadora digital.

§. El argumento de la informalidad de comportamiento

No se puede elaborar un conjunto de reglas para describir lo que una persona hace en todas las circunstancias concebibles. Puede establecerse la regla de que, por ejemplo, hay que detenerse al ver

un semáforo rojo y continuar si se ve uno verde, pero ¿qué sucede si, por un error, se iluminan los dos a la vez? Quizá la persona decida que es mejor detenerse. Pero por esta decisión pueden surgir ulteriormente dificultades. Intentar sentar reglas de conducta que cubran cualquier eventualidad, hasta las resultantes de las luces de tráfico, parece imposible. Estoy de acuerdo con esto.

A partir de ello se arguye que no podemos ser máquinas. Trataré de exponer el argumento, pero temo no hacerle debidamente justicia. Al parecer, se desarrolla de este modo: «Si cada persona posee un conjunto fijo de reglas de conducta por las que rige su vida, no sería más que una máquina; pero no hay tales reglas. Por lo tanto, las personas no pueden ser máquinas». Es deslumbrante el injusto medio. No creo que el argumento se plantee casi nunca así, pero estoy convencido de que constituye la base de la argumentación. Sin embargo, puede darse cierta confusión entre «reglas de conducta» y «leyes de comportamiento» para oscurecer la conclusión. Por «reglas de conducta» entiendo preceptos tales como «Pare si ve luces rojas» que uno puede cumplir conscientemente. Por «leyes de comportamiento» entiendo leyes naturales aplicables al cuerpo humano, tales como «si le pellizcas, chilla». Si sustituimos «leyes de comportamiento que regulan su vida» por «leyes de conducta por las que rige su vida», el injusto medio deja de ser insuperable en el argumento en cuestión, pues creemos que no sólo es cierto que estar regulado por leyes de comportamiento implica ser una especie de máquina (aunque no necesariamente una máquina de estado discreto), sino que, a la inversa, ser tal máquina implica estar

regulado por tales leyes. Sin embargo, no podemos convencernos tan fácilmente de la ausencia total de leyes de comportamiento como de la ausencia absoluta de leyes de conducta. El único modo de descubrir tales leyes consiste en la observación científica, y no conocemos circunstancias en las que pueda decirse: «Ya hemos buscado bastante. No existen tales leyes».

Podemos demostrar más categóricamente que semejante afirmación es injustificada. Supongamos que fuera posible con absoluta seguridad descubrir esas leyes, si existiesen. Entonces, dada una máquina de estado discreto, no cabe duda de que podría descubrirse, mediante la observación suficiente para predecirlas, su comportamiento futuro, y eso dentro de un tiempo razonable, digamos mil años. Pero no parece ser el caso. He elaborado en la computadora de Manchester un pequeño programa con tan sólo 1000 unidades de almacenamiento, merced al cual, si se entrega a la máquina una cifra de dieciséis guarismos, responde con otra de igual magnitud en dos segundos. Desafío a cualquiera a que descubra en esas respuestas suficientes datos sobre el programa para ser capaz de predecir cualquier respuesta a valores no probados.

§. El argumento de la percepción extra-sensorial

Supongo que el lector está al corriente de la idea de percepción extra-sensorial y del significado de sus cuatro variantes: telepatía, clarividencia, precognición y psicocinesis. Estos extraños fenómenos parecen refutar todas las ideas cinéticas habituales, ¡Cuánto nos

gustaría desacreditarlos! Pero lamentablemente la evidencia estadística, al menos en el caso de la telepatía, es abrumadora. Resulta difícil para cualquiera reajustar sus propias ideas para dar cabida a estos hechos singulares, pero, una vez admitidos, no parece que cueste mucho creer en fantasmas y espíritus. Lo primero que se nos ocurre es la idea de que nuestros cuerpos se mueven de modo simple con arreglo a las leyes físicas conocidas, junto a otras no descubiertas pero bastante parecidas.

Para mí es un argumento de bastante peso. Podría argüirse que muchas teorías científicas siguen siendo válidas en la práctica, a pesar de que contradigan la percepción extra-sensorial, y que puede prescindirse perfectamente de ella, pero no deja de ser un conformismo fácil; precisamente es muy de temer que no sea el pensamiento el tipo de fenómeno en el que la percepción extra-sensorial sea particularmente relevante.

Un argumento más específico basado en la percepción extra-sensorial sería el siguiente: «Juguemos al juego de imitación, teniendo por testigo a una persona que sea buena receptora telepática y a una computadora digital. El preguntador puede plantear preguntas de este tipo: "¿A qué palo pertenece la carta que tengo en mi mano derecha?". La persona, mediante telepatía o clarividencia, da la respuesta correcta 130 veces sobre 400 cartas. La máquina sólo puede adivinar al azar y tal vez acierte 104 veces, y así el preguntador efectúa la identificación correcta». Esta es una interesante posibilidad. Supongamos que la computadora digital cuenta con un generador numérico aleatorio, es natural que lo

utilice para dar la respuesta. Pero el generador numérico aleatorio está sujeto al poder psicocinético del preguntador y quizás esta psicocinesis sea la causa de que la máquina acierte más veces de las que cabría esperarse de un cálculo de probabilidades, por lo que el preguntador seguiría siendo incapaz de efectuar la identificación correcta. Por otra parte, puede ser capaz de acertar sin plantear preguntas, gracias a la clarividencia. Con la percepción extra-sensorial puede suceder cualquier cosa.

Si admitimos la telepatía, habrá que depurar la prueba. Puede considerarse la situación similar a la que se produce si el preguntador hablara consigo mismo y uno de los participantes estuviera escuchando con el oído en la pared. Situando a los participantes en una «habitación a prueba de telepatía», se restablecerían las condiciones.

Capítulo 7
Máquinas que aprenden

Habrá comprobado el lector que no dispongo de argumento positivo alguno lo bastante convincente para apoyar mi tesis. Si lo tuviera, no me habría tomado tanta molestia en exponer detalladamente las falacias de las tesis contrarias. Ahora expondré la evidencia en favor de mi punto de vista.

Volvamos brevemente a la objeción de lady Lovelace, quien afirmaba que la máquina sólo puede hacer lo que nosotros le mandemos. Podríamos decir que una persona puede «inyectar» una idea en una máquina y que ésta respondería hasta cierto límite, quedándose quieta a continuación, como la cuerda de un piano percutida por un martillo. Otro símil podría ser una pila atómica de tamaño inferior al crítico: una idea inyectada correspondería a un neutrón que penetra desde fuera en la pila. Cada uno de estos neutrones provoca una determinada alteración que acaba por disiparse. Sin embargo, si aumentamos suficientemente el tamaño de la pila, la alteración causada por el neutrón incluso irá en aumento hasta la completa destrucción de la pila. ¿Existe un fenómeno equivalente para las mentes, y se da también en el caso de las máquinas? En el caso de la mente humana parece haberlo. La mayoría de los cerebros parecen ser «subcríticos», es decir, que corresponden en esta analogía a pilas de tamaño subcrítico. Una idea presentada a este tipo de mente, no inducirá generalmente más que una idea por respuesta. Una reducidísima proporción de cerebros son

supercríticos. En ellos una idea da origen a toda una «teoría» formada por ideas secundarias, terciarias y de todo orden. Las mentes animales parecen decididamente ser subcríticas. Siguiendo la analogía, nos preguntamos: « ¿Se puede hacer que una máquina sea supercrítica?».

La analogía de la «piel de cebolla» también es válida. Si consideramos las funciones de la mente o del cerebro, observamos determinadas operaciones explicables en términos puramente mecánicos. Lo que decimos no es aplicable a la auténtica mente: es una especie de piel que hay que quitar si queremos verla realmente. Pero, luego, en lo que queda, encontramos otra piel que hay que eliminar, y así sucesivamente. Con este método, ¿llegamos con seguridad a la mente «real», o simplemente a la piel que no encierra nada? En tal caso toda mente es mecánica. (De todas formas, hemos explicado ya que no es una máquina de estado discreto).

Los últimos párrafos no pretenden ser argumentos convincentes, sino más bien deben tomarse como «una letanía destinada a inculcar una creencia».

El único apoyo realmente satisfactorio que puede darse a la opinión manifestada al principio del apartado 6 es el que consiste en esperar a finales de siglo y luego efectuar el experimento señalado. ¿Pero qué podemos decir entretanto? ¿Qué pasos hemos de dar ahora para que dé buen resultado el experimento?

Como he dicho, el problema fundamental estriba en programar. También serán imprescindibles progresos de ingeniería, pero creo que estarán a la altura de las necesidades. Las estimaciones de la

capacidad de almacenamiento del cerebro oscilan entre 10^{10} y 10^{15}dígitos binarios. Personalmente me inclino por el valor más bajo y creo que sólo una pequeña parte se utiliza para los tipos más elevados de pensamiento. La mayor parte de esta capacidad se emplea seguramente en la retención de impresiones visuales. Me sorprendería que se necesitara más de 10^9 para jugar bien al juego de imitación, en cualquier caso contra un hombre ciego. (Nota: la capacidad de la *Encyclopaedia Britannica*, decimoprimera edición, es de 2×10^9). Una capacidad de almacenamiento de 10^7 sería una posibilidad bastante real, aun con las técnicas actuales. Probablemente no será preciso aumentar la velocidad de operación de las máquinas. Partes de las máquinas modernas, que podríamos calificar de auténticas células nerviosas, trabajan mil veces más rápido que éstas. Con esto se conseguiría un «margen de seguridad» para compensar pérdidas de velocidad producidas por diversos motivos. El problema estriba, en último extremo, en saber cómo programar estas máquinas para jugar al juego. Con mi actual ritmo de trabajo produzco unos mil dígitos de programa diarios; en consecuencia, unas sesenta personas, trabajando asiduamente durante cincuenta años, podrían llevar a cabo esta tarea si no traspapelaran nada. Parece deseable un método más expeditivo.

En el proceso de intentar la imitación de una mente humana adulta estamos obligados a pensar muy en serio sobre el proceso por el que se ha llegado al estado en que se halla. Se observarán tres factores:

1. *El estado inicial de la mente al nacer.*
2. *La educación que ha tenido.*

3. Otras experiencias, aparte de la educación, a que haya estado sometida.

En lugar de intentar la elaboración de un programa que imite la mente adulta, ¿por qué no establecer uno que simule la mente infantil? Si luego la sometemos a un curso adecuado de formación, podría obtenerse un cerebro adulto. Podemos decir que el cerebro infantil es como el cuaderno recién comprado en una papelería: poco mecanismo y muchas hojas en blanco. (Mecanismo y escritura son casi sinónimos desde nuestro punto de vista). Nuestra esperanza se funda en que hay tan poco mecanismo en el cerebro infantil que debe resultar fácil programar algo similar. Podemos suponer que la cantidad de trabajo formativo, en una primera aproximación, sea muy parecida a la aplicable en el caso de un niño.

De este modo, el problema queda dividido en dos partes: el programa infantil y el proceso formativo. Ambos estrechamente vinculados. No puede esperarse construir una buena máquina infantil al primer intento; hay que experimentar enseñando a la máquina, y comprobar si aprende bien. Luego puede probarse otra vez y ver si es mejor o peor. Evidentemente existe una clara relación por analogía entre este proceso y el de la evolución:

Estructura de la máquina infantil = Material hereditario

Cambios de la máquina infantil = Mutaciones

Selección natural = Juicio del experimentador

Sin embargo, es de esperar que este proceso sea más expeditivo que el de la evolución. La supervivencia del más apto es un método lento para valorar las ventajas. El experimentador, aplicando su inteligencia, debe ser capaz de acelerarlo. De igual importancia es el hecho de que no está limitado por mutaciones aleatorias. Si el experimentador descubre la causa de determinada debilidad, puede probablemente decidir el tipo de mutación que la mejore.

A la máquina no se le podrá aplicar exactamente el mismo proceso de aprendizaje que a un niño. Ya que, por ejemplo, no tendrá piernas y no se le podrá ordenar que vaya a por un cubo de carbón. Seguramente tampoco tendrá ojos. Y por mucho que se compensen estas deficiencias con una buena ingeniería, no se podrá enviar a la criatura a la escuela porque sería motivo de burla de sus compañeros. Habrá que darle clases particulares, sin preocuparnos por las piernas, los ojos, etc. El caso de Helen Keller demuestra que es posible la labor educativa a condición de que se establezca una comunicación bilateral entre maestro y alumno por el medio que sea.

Normalmente asociamos castigos y recompensas al proceso educativo. Algunas máquinas infantiles simples pueden construirse o programarse ateniéndose a ese principio. Hay que construir la máquina de tal modo que los acontecimientos que preceden brevemente a la aparición de la señal de castigo cuenten con mínimas posibilidades de repetición, y que, por el contrario, la señal de recompensa incremente la posibilidad de repetición de secuencias que la motivan. Estas especificaciones no presuponen

tipo de sentimiento alguno por parte de la máquina. He realizado algunos experimentos con este tipo de máquina infantil y he logrado enseñarle varias cosas, pero utilicé un método de aprendizaje excesivamente heterodoxo para que el experimento pueda considerarse un éxito.

El empleo de castigos y recompensas puede a lo sumo formar parte del proceso de aprendizaje. En términos generales, si el enseñante no dispone de otros medios de comunicación con el alumno, la cantidad de información que éste recibe nunca excede el número de recompensas y castigos. Cuando un niño ha aprendido finalmente a repetir «Casabianca», se sentirá probablemente muy afligido si la única manera de dilucidar el texto es la técnica de las «Veinte preguntas» y cada «NO» supone una bofetada. Por lo tanto, es necesario disponer de otros canales de comunicación «no emocionales». Si los hay, se puede enseñar a una máquina por el método de premios y castigos a obedecer órdenes dadas en una lengua determinada, es decir un lenguaje simbólico. Estas órdenes se transmiten por canales «no emocionales», y el empleo de dicho lenguaje disminuye notablemente la cantidad de castigos y premios.

Puede existir diversidad de opiniones en cuanto a la complejidad adecuada de la máquina infantil. Puede intentarse una construcción lo más simple posible, coherente con los principios generales. O puede dotársela de un sistema completo integrado de inferencia lógica, en cuyo caso el almacenamiento estará fundamentalmente ocupado por definiciones y proposiciones. Estas proposiciones serían de diversa índole: hechos bien establecidos, conjeturas,

teoremas matemáticamente demostrables, afirmaciones hechas por una autoridad, expresiones con forma lógica de proposición pero de valor no creíble. Algunas proposiciones serían «imperativas». La máquina estaría construida de forma que, en cuanto una imperativa se clasificara como «bien establecida», se produjera automáticamente la acción apropiada. Como ejemplo, supongamos que el maestro dice a la máquina: «Ahora haz los deberes». Esto podría dar lugar a que «El maestro dice "Ahora haz los deberes"» quedara incluido en los hechos bien establecidos. Otra posibilidad sería: «Todo lo que dice el maestro es cierto». Ambas posibilidades combinadas podrían dar por resultado que la imperativa «Ahora haz los deberes» quedara incluida entre los hechos bien establecidos, lo cual, con arreglo a la construcción de la máquina, significaría que se inician realmente los deberes, pero el efecto es muy poco satisfactorio. El proceso de inferencia que utilice la máquina tiene que satisfacer al lógico más riguroso. Por ejemplo, no habrá jerarquía de tipos, lo que no significa que no se produzcan falacias de tipos, semejantes al riesgo de caer por un precipicio no señalizado. Unos imperativos adecuados (expresados *dentro* de los sistemas, pero que no formen parte de las reglas *del* sistema), tales como «No emplees una clase si no es una subclase de las mencionadas por el maestro», ejercerían la misma función que un letrero que indicara: «No acercarse al borde».

Las imperativas a las que obedece una máquina sin miembros son necesariamente de índole intelectual, como en el ejemplo citado (hacer los deberes). Entre dichas imperativas son importantes las

que rigen el orden en que hay que aplicar las reglas del sistema lógico correspondiente, ya que, en cada fase de la utilización de un sistema lógico, hay una amplia alternativa de pasos que pueden seguirse para no transgredir las reglas de ese sistema lógico. Estas opciones marcan la diferencia entre un razonador brillante y otro torpe, pero no la diferencia entre uno serio y otro tramposo. Las proposiciones que conducen a las imperativas de esta clase pueden ser: «Cuando se mencione a Sócrates, utiliza el silogismo en Bárbara», o «Si se ha demostrado que un método es más rápido que otro, no uses el método lento». Algunas pueden «basarse en una autoridad», pero otras puede producirlas la propia máquina por inducción científica, por ejemplo.

La idea de una máquina que aprende puede parecer paradójica a algunos lectores. ¿Cómo pueden cambiarse las reglas de operación de la máquina? Estas deben especificar punto por punto cómo debe reaccionar la máquina independientemente de su historia y al margen de los cambios que experimente. Por lo tanto, las reglas son bastante invariables con respecto al tiempo. Y es bien cierto. La explicación de la paradoja consiste en que las reglas que cambian en el proceso de aprendizaje son de un tipo menos pretencioso y sólo tienen validez efímera. El lector puede establecer un paralelismo con la Constitución de los Estados Unidos.

Una característica importante de la máquina que aprende es la de que el profesor ignora muchas veces la mayoría de los procesos internos, aunque hasta cierto punto sea capaz de predecir el comportamiento de su alumno. Esto es tanto más aplicable a la

formación ulterior de una máquina que tenga por origen una máquina infantil con un diseño (o programa) perfectamente experimentado. Situación muy distinta al procedimiento normal de emplear una máquina para hacer cálculos, ya que el objeto, en este caso, consiste en disponer de una imagen mental clara del estado de la máquina en cada momento de la computación. Este propósito sólo es alcanzable con una imposición. La opinión de que «la máquina sólo hace lo que queramos que haga parece extraña a la vista de lo expuesto». La mayoría de los programas que podemos introducir en la máquina la hará hacer algo que no entendemos o que consideramos como comportamiento totalmente aleatorio. El comportamiento inteligente consiste probablemente en una desviación del comportamiento absolutamente disciplinado que implica la computación, aunque relativamente leve y sin que provoque un comportamiento aleatorio o loops repetitivos inútiles. Otro importante resultado de la preparación de una máquina para que intervenga en el juego de imitación, merced a un proceso de enseñanza y aprendizaje, radica en que la «falibilidad humana» suele quedar descartada de una forma bastante natural, sin necesidad de «entrenamiento» especial. Los procesos que se aprenden no procuran una certeza absoluta de resultados; si así fuera, nunca fallaría su aprendizaje.

Quizá convenga introducir un elemento aleatorio en la máquina que aprende. Un elemento aleatorio resulta bastante útil en la búsqueda de la solución de un problema. Supongamos, por ejemplo, que deseamos hallar un número entre 50 y 200 que sea igual al

cuadrado de la suma de sus cifras; empecemos por el 51 y sigamos con el 52 hasta encontrar la combinación justa. Otra alternativa sería elegir números al azar hasta hallar uno que nos sirva. Este método presenta la ventaja de que nos ahorra la necesidad de mantener el registro de los valores que se han probado, y el inconveniente de que se corre el riesgo de probar dos veces el mismo número, pero esto no es tan importante si hay varias soluciones. El método sistemático presenta el inconveniente de que puede haber una serie enorme sin solución en la región que hay que investigar en primer lugar. El proceso de aprendizaje puede considerarse como la búsqueda de una forma de comportamiento que satisfaga al profesor (o cualquier otro requisito). Como probablemente existe un gran número de soluciones satisfactorias, el método aleatorio parece mejor que el sistemático. Se advertirá que es el que interviene en el proceso análogo de la evolución, y que en ella no es posible el método sistemático. ¿Cómo sería posible conservar el registro de las distintas combinaciones genéticas ensayadas para evitar probarlas de nuevo?

Esperemos que las máquinas lleguen a competir con el hombre en todos los campos puramente intelectuales. ¿Pero cuáles son los mejores para empezar? También es una ardua decisión. Muchos piensan que lo mejor es una actividad de naturaleza tan abstracta como jugar al ajedrez. También puede sostenerse que lo óptimo sería dotar a la máquina de los mejores órganos sensoriales posibles y luego enseñarla a entender y a hablar inglés. Es un proceso que podría hacerse con arreglo al aprendizaje normal de un niño: se

señalan los objetos, se los nombra, etc. Vuelvo a insistir en que ignoro la respuesta adecuada; creo que hay que experimentar los dos enfoques.

Sólo podemos prever el futuro inmediato, pero de lo que no cabe duda es que hay mucho por hacer.

Autor

ALAN MATHISON TURING (Paddington, Londres, 23 de junio de 1912 - Wilmslow, Cheshire, 7 de junio de 1954). Fue un matemático, lógico, científico de la computación, criptógrafo y filósofo británico.

Es considerado uno de los padres de la ciencia de la computación siendo el precursor de la informática moderna. Proporcionó una influyente formalización de los conceptos de algoritmo y computación: la máquina de Turing. Formuló su propia versión de la hoy ampliamente aceptada Tesis de Church-Turing.

Nació en Londres (Gran Bretaña), desde muy temprana edad Turing demostró su inteligencia. A los 3 años tenía una inusual capacidad para recordar palabras y a los 8 años se interesó por la química montando un laboratorio en su casa. Con 13 años ingresó en la

escuela Sherborne, en la que ya demostraba su facilidad para las matemáticas, teniendo una gran capacidad para realizar cálculos mentalmente.

Obtuvo una beca para estudiar en la universidad de Cambridge, en donde se graduó de la licenciatura de matemáticas con honores en 1934. En abril de 1936, publicó el artículo «*On computable numbers, with an application to the Entscheidungs problem*» en el que introduce el concepto de algoritmo y de máquina de Turing. Este artículo da respuesta (negativa) al problema de la decisión formulada por Hilbert en 1900, probando que existen problemas sin solución algorítmica y es uno de los cimientos más importantes de la teoría de la computación.

En septiembre de 1936, Turing ingresó en la universidad de Princeton (EE.UU.). Su artículo atrajo la atención de uno de los científicos más destacados de la época, John von Neumann, quien le ofreció una beca en el Instituto de Estudios Avanzados. Turing obtuvo su doctorado en matemáticas en 1938. Tras su graduación, von Neumann le ofreció una plaza como su asistente, pero Turing rechazó la oferta y volvió a Inglaterra, en donde vivió de una beca universitaria mientras estudiaba filosofía de las matemáticas entre 1938 y 1939.

En 1939, con el comienzo de la Segunda Guerra Mundial, Turing fue reclutado por el ejército británico para descifrar los códigos emitidos por la máquina Enigma utilizada por los alemanes. En el deseo de obtener mejores máquinas descifradoras, se comenzó a construir la primera computadora electrónica, llamada Colossus,

bajo la supervisión de Turing, se construyeron 10 unidades, y la primera empezó a operar en 1943. Por su trabajo en el Colossus, Turing recibió la Orden del Imperio Británico en 1946.

En 1944, Turing fue contratado por el Laboratorio Nacional de Física (NLP) para competir con el proyecto americano EDVAC, de von Neumann. Turing ejerció como Oficial Científico Principal a cargo del Automatic Computing Engine (ACE). Hacia 1947, Turing concibió la idea de las redes de cómputo y el concepto de subrutina y biblioteca de software. También describió las ideas básicas de lo que hoy se conoce como red neuronal. Abandonó la NLP en 1948.

Turing se adelantó al proyecto de construcción de un ordenador de acuerdo con la arquitectura de von Neumann. El Manchester Mark I, estuvo acabado en 1948 antes que el EDVAC. Turing diseñó para esta máquina un lenguaje de programación basado en el código empleado por los teletipos.

Otro de los campos de investigación de Turing fue la inteligencia artificial, se puede decir que esta disciplina nació a partir del artículo titulado «*Computing Machinery and Inteligence*» publicado por Turing en 1950. Es muy famosa la primera frase de este artículo: «Propongo considerar la siguiente cuestión: ¿Pueden pensar las máquinas?». Turing propuso un método llamado el test de Turing para determinar si las máquinas podrían tener la capacidad de pensar.

En 1951, es nombrado miembro de la Sociedad Real de Londres por sus contribuciones científicas. Y en su honor, la Association for Computing Machinery llama «Turing Award» a su premio más

importante, el cual se otorga desde 1966 a los expertos que han realizado las mayores contribuciones al avance de la computación.

La carrera de Turing terminó súbitamente después de ser procesado por ser homosexual. Turing se suicidó dos años después de su condena.

El 24 de diciembre de 2013, la reina Isabel II de Inglaterra promulgó el edicto por el que se exoneró oficialmente al matemático, quedando anulados todos los cargos en su contra.