

ALAN TURING: BASES, FORMA Y CRÍTICAS A LA INTELIGENCIA ARTIFICIAL

ALAN TURING: BASE, FORM AND CRITICISM
TO ARTIFICIAL INTELLIGENCE

GUSTAVO ESPARZA

Doctorado en Ciencias Sociales y Humanidades
Profesor Investigador
Universidad Panamericana
Aguascalientes/México
gaesparza@up.edu.mx
ORCID: 0000-0002-9470-6519

Recibido: 9/04/2021
Revisado: 6/08/2021
Aceptado: 6/09/2021

Resumen: El objetivo del presente artículo es definir los elementos que aportó Alan Turing al desarrollo del concepto de *Inteligencia Artificial (AI)*. Como problema se propone reflexionar las aportaciones del matemático a la génesis del término como lo conocemos hoy en día. Metodológicamente se estudian las obras centrales del autor para delimitar los fundamentos lógicos y epistemológicos a través de los cuales opera una *Computadora Inteligente*, para con ello valorar las aportaciones a la disciplina computacional y la eventual formulación del término. La presente investigación ofrece tres resultados: (1) Turing concibe a las computadoras como auténticos objetos materiales pensantes; (2) el modelo de diseño de una computadora con estas características, de acuerdo al autor, debe considerar el patrón de desarrollo cerebral (y aprendizaje) infantil; (3) el debate póstumo sobre el pensamiento de Turing fue decisivo para la construcción de la agenda de problemas que dio origen al término *AI*.

Palabras Clave: Computación, Matemática, Filosofía, Inteligencia Artificial, Turing.

Abstract: The objective of this article is to define the elements that Alan Turing contributed to the development of the concept of Artificial Intelligence (AI). As a problem, it is proposed to study the contributions of the mathematician to the genesis of the term as we know it today. Methodologically, the central works of the author are studied to define the logical and epistemological foundations through which an Intelligent Computer operates, in order to assess the contributions to the computational discipline and the eventual formulation of the term. The present investigation offers three results: (1) Turing conceives computers

as authentic thinking material objects; (2) the design model of a computer with these characteristics, according to the author, must consider the pattern of brain development (and learning) in children; (3) the posthumous debate on Turing's thought was decisive for the construction of the agenda of problems that gave rise to the term *AI*.

Keywords: Computation, Mathematics, Philosophy, Artificial Intelligence, Turing.

1. INTRODUCCIÓN. INTELIGENCIA NATURAL Y ARTIFICIAL¹

El objetivo del siguiente artículo es estudiar los fundamentos y concepción de una Computadora Inteligente propuesta por Alan Turing y las aportaciones que se derivaron de esta propuesta al desarrollo de una *Artificial Intelligence (AI)*; el interés es mostrar tres fases cruciales en la configuración de una máquina pensante: primera, la creación de un mecanismo que fuese capaz de resolver si una fórmula de primer orden es lógicamente válida; segunda, la apología que realizó el matemático inglés sobre la posibilidad de fundamentar que las computadoras fuesen objetos inteligentes; tercera, el debate que se generó en torno estas formulaciones previas y que contribuyeron a la construcción de una agenda de problemas filosóficos y computacionales relacionados con el surgimiento de una *AI*².

Es importante mencionar que Alan Turing no ofreció una definición de *AI*, pero sus contribuciones fueron decisivas para el posterior surgimiento del término. Como tal, fue John McCarthy y un grupo de investigadores quienes durante un encuentro en Darmouth en 1956, propusieron la siguiente agenda de objetivos como parte de las tareas a desarrollar:

The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, from abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves³.

1 Agradezco la lectura y comentarios de Daniel Martínez-Barba, Belen Prado y de los revisores anónimos por su lectura y comentarios a versiones previas de este manuscrito. Sus aportaciones fueron valiosas para la precisión de la tesis central de esta investigación. El presente artículo se encuentra bajo el auspicio del Fondo Fomento a la Investigación 2020, con código: UP-CI-2020-MEX-13-FIL: Lógica, Epistemología y Ética de la Inteligencia Artificial.

2 TURING, Alan. *The essential Turing. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secret of Enigma*, (Edición de Jack Copeland. Oxford: Oxford University Press, 2004). Esta edición de Copeland reúne las obras más importantes del autor y, además, ofrece estudios críticos que aclaran la importancia de los trabajos del matemático inglés; en adelante: *The essential Turing*.

3 MCCARTHY, John; MINSKY, Marvin; ROCHESTER, Nathaniel; SHANNON, Claude. *A proposal for the Darmouth Summer Research Project on Artificial Intelligence*, 1955, p. 1.

De este planteamiento se propusieron siete tareas específicas, tales eran: (1) *Computadoras automáticas*: la velocidad y la memoria de las computadoras de ese tiempo era insuficiente para simular las funciones superiores del cerebro humano, sin embargo, consideraban que el problema central estaba en la (in)capacidad del programador para escribir los programas. (2) *¿Cómo programar una computadora para que use lenguaje?*: al asumir que el cerebro humano utiliza y codifica signos lingüísticos, estudiar el modo en que la generalización de nuevas palabras implica el seguimiento de reglas previas y nuevas. (3) *Redes neuronales*: buscar el modo en que un grupo de neuronas se organizan para formar un concepto. (4) *Tamaño del cálculo*: desarrollar criterios eficientes para medir modos eficientes para resolver un problema. (5) *Mejora autónoma*: encontrar los medios a través de las cuales una computadora inteligente busca cambios automáticamente de modo independiente. (6) *Abstracciones*: describir y precisar los procesos mediante los cuales se clarifique el proceso de abstracción. (7) *Aleatoriedad y creatividad*: estudiar los mecanismos mediante los cuales la aleatoriedad se vincula a un pensamiento intuitivo que establece un principio de orden.

John McCarthy, en una serie de ensayos recolectados posteriormente, reconocería que ese encuentro de Dartmouth en 1956 sería considerado decisivo para la génesis del concepto, pero sobre todo para el desarrollo de una agenda de investigación cuyo marco de objetivos partiera de estas siete tareas. En la retrospectiva sobre los logros, el autor menciona que la actividad central del encuentro fue el estudio de “las relaciones entre el cerebro y la computadora”⁴, cuestión que tuvo como referencia los trabajos de Turing: “[he] proposed that mental processes be studied by programming a computer to carry them out rather than by building machines that imitate the brain at the physiological level”⁵. De esta manera, para los iniciadores del término ‘AI’ la visión del matemático inglés fue determinante para la construcción de la agenda de problemas ya mencionados.

El presente artículo estudia el desarrollo del pensamiento de Alan Turing, así como sus aportaciones al campo de la computación. En el progreso de las investigaciones del matemático inglés, se aprecian referencias a conceptos y fundamentaciones sobre “motores de cálculo automático”, “máquinas pensantes”, “computadoras inteligentes” y “computadoras digitales pensantes”, pero no propiamente el término *inteligencia artificial*⁶, aún con ello, la reconstrucción de los

Disponible en línea: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> (última consulta 2 de agosto de 2021).

4 McCarthy, John, *Defending AI Research. A collection of Essays and Reviews* (Stanford: CSLI Publications, 1996), p. 45.

5 McCarthy, John, *Defending AI Research*, p. 46.

6 Los detalles y fundamentos de cada uno de los términos, se pueden consultar en: *The essential Turing*, pp. 362-506, más adelante revisaremos las definiciones y argumentos aducidos para

resultados que llevaron a preguntar *Can Machines think?* fueron decisivos para que McCarthy y su grupo logaran sus objetivos en el Dartmouth College⁷.

En el resto del artículo, se expone el pensamiento computacional de Turing desde el contexto histórico que animó a imaginar una solución al problema matemático de la completitud, en donde su primera propuesta concebía un procesamiento de la información del tipo “sí... entonces”, hasta sus reflexiones sobre las posibilidades de construir una computadora inteligente, cuyas características presentan una primera aproximación a lo que hoy se conoce como “aprendizaje por refuerzo”⁸; el conjunto permitirá delimitar los aportes que el grupo de Dartmouth tuvo en cuenta para el desarrollo del término *AI*.

2. MATEMÁTICA Y COMPUTACIÓN. LA VALIDACIÓN DEL TEOREMA DE LA COMPLETITUD

Con respecto a los antecedentes del concepto *AI* es necesario remitirse a una conferencia dictada en 1900 en París por David Hilbert, en donde delimitaba una serie de problemas matemáticos que debían ser resueltos por esta ciencia; más tarde, él mismo, en 1927 durante el Seminario de Matemáticas de Hamburg, pronunciaría una conferencia titulada *The Foundations of Mathematics* en la cual sostendría lo siguiente:

The source of pure existence theorems is the logical c-axiom, upon which in turn the constructions of all ideal propositions depends... This formula game enables us to express the entire thought-content of the science of mathematics in a uniform manner and develop it in such a way that, in a same time, the interconnections between the individual propositions and facts become clear⁹.

La aspiración de Hilbert, si bien se colocaba dentro de la esfera puramente matemática, requería de la lógica para estructurar los fundamentos. De acuerdo a sus postulados había que probar que los axiomas no eran contradictorios entre sí, para lo cual había que demostrar que partiendo de éstos no se llegaría nunca

cada uno de estos términos.

7 Aunque en la convocatoria del encuentro no se menciona como antecedente la obra de Turing, años después el mismo McCarthy reconocería que el trabajo “Computing Machinery and Intelligence” (*The essential Turing*, pp. 441-464) fue una referencia central para situar la génesis de problemas de una *AI*.

8 Agradezco a Belen Prado por hacerme notar esta distinción en el pensamiento de Turing.

9 HILBERT, David. “On the Foundations of mathematics”, (en: VAN HEIJENOORT, Jean, *From Frege to Gödel. A Source Book in Mathematical Logic, 1879-1931*. Cambridge: Harvad University Press, pp. 464-479, 1967), p. 495.

a resultados contrarios siguiendo un número finito de deducciones lógicas¹⁰. Para probar la veracidad de esta cuestión, había que solventar, primero, lo que el matemático alemán llamaría *Entscheidungsproblem* (problema de la decisión)¹¹. Otro matemático inglés, Alonzo Church, definiría este dilema del modo siguiente:

By the *Entscheidungsproblem* of a system of symbolic logic is here understood the problem to find an effective method by which, given any expression Q in the notation of the system, it can be determined whether or not Q is provable in the system¹².

De acuerdo a esta definición, era comprensible que los estudiosos de la época buscaran demostrar el postulado de Hilbert; tres serían las más importantes refutaciones a este proyecto: la de Gödel, la de Church y la de Turing. Primero, en 1931 Kurt Gödel afirmaría que no era posible determinar un axioma que estableciera la veracidad o falsedad de una proposición aritmética, pues existían limitaciones para tales efectos: “In any of the formal systems... there are undecidable problems of the restricted functional calculus (that is, formulas of the restricted functional calculus for which neither validity nor the existence of a counterexample is probable)”¹³. Por su parte, Church, desarrollando el cálculo lambda, demostraría que era imposible determinar si una fórmula bien formada era o no un teorema: “no effective decision procedure exists which suffices to determine of an arbitrary [well formed formula] whether or not it is a theorem”¹⁴. En general, tanto Gödel como Church explicaban que dada la afirmación primera de un teorema **L** como falso, podía admitirse, en una segunda instancia, una sustitución de sus postulados iniciales por otros que fueran verdaderos, pero dado que la proposición inicial había sido demostrada como falsa o irresoluble, al sustituirse por **L'** (cuya redefinición permitía resolver el teorema cumplidas las nuevas condiciones) entonces se implicaría que **L=L'**; en este nuevo contexto, siendo **L** verdadera o siendo plausible su solución, se derivaba un conflicto con la primera afirmación en

10 HILBERT, David, *Mathematical Problems* (ICM París, 1990): “But above all I wish to designate the following as the most important among the numerous questions which can be asked with regard to the axioms: *To prove that they are not contradictory, that is, that a definite number of logical steps based upon them can never lead to contradictory results*”, p. 8. Énfasis en el original.

11 Recomiendo ampliamente el trabajo de Belen Prado (“Ambivalence in Machine Intelligence: The epistemological Roots of the Turing Machine”, *Signos Filosóficos*, Vol. 23, No. 45 (2021), pp. 57-61) para valorar las implicaciones epistemológicas de este problema.

12 CHURCH, Alonzo. “A Note on *Entscheidungsproblem*”, (*The Journal of Symbolic Logic*, Vol. 1, No.1 (1936), pp. 40-41), p. 41, nota 6.

13 GÖDEL, Kurt. “On formally undecidable propositions of *Principia Mathematica* and related systems I”, (en: Kurt Gödel, *Collected Works, Vol. I (Publications 1929-1936)*. Edited by Solomon Feferman. Oxford: Oxford University Press, pp. 144-195, 1986), p. 187. En adelante *Kurt Gödel, I*.

14 CHURCH, Alonzo. *Introduction to Mathematical Logic. Vol. I*, (New Jersey: Princeton University Press, 1956), p. 246.

la que **L** era falsa. Finalmente Turing¹⁵, propondría una demostración del teorema de Hilbert recurriendo a un experimento computacional, concluyendo que, en aquellos casos en las que una proposición **L** falsa, al sustituirse por **L'** verdadera, dicha proposición se traduciría en la circularidad de la máquina generando que esta trabajase indefinidamente. Ello se interpretaría como la imposibilidad para demostrar el problema de la decisión: “the Hilbertian *Entscheidungsproblem* [have] no solution”¹⁶. De las tres aportaciones, en lo que resta de la sección, únicamente se retomará la ruta computacional considerada.

En el artículo de Turing, “On Computable Numbers, with an Application to the *Entscheidungsproblem*”¹⁷, la demostración comenzaba asumiendo la existencia de una máquina que tenía la siguiente fórmula de programación:

$$\delta: Q \times \Gamma \rightarrow \Gamma \times \{L, R\} \times Q^{18}$$

En donde:

δ : es la función de transición.

Q : es el estado en el que se encuentra la máquina ($q_1, q_2, q_3 \dots q_n$)

Γ : es el alfabeto leído por la máquina, el cual puede ser ($\square, 0, 1$)

L, R : la dirección que puede tomar la máquina en la cinta o memoria (izquierda o derecha).

Un ejemplo sería:

$$\delta(q_0, 0) = (0, R, q_1)$$

[ubicada la máquina en un estado inicial q_0 , lee 0, deja 0, y pasa a la derecha de la cinta para ubicarse en el estado q_1].

La explicación de Turing¹⁹ sobre el funcionamiento de la máquina (inicialmente presentada como parte de un experimento mental), concebía algunas alternativas de funcionamiento en las que se posibilitaba la impresión de números o “Digital

15 TURING, Alan, *The essential Turing*, pp. 58-90.

16 TURING, Alan, *The essential Turing*, p. 59.

17 TURING, Alan, *The essential Turing*, pp. 58-90.

18 Para la elaboración de estos ejemplos me he guiado de la exposición de Javier García (*La máquina de Turing (explicada)*). Conferencia 2016, en: <https://www.youtube.com/watch?v=NS-NQ5mCSs8>, pues aprovecha una simbología hoy en día mayoritariamente aceptada.

19 Turing (*The essential Turing*, pp. 59-68) ofrece una explicación más amplia de estas ideas, pero usando otros símbolos y apoyándose de los números binarios. Un ejemplo que él propone es el siguiente:

“In this way we reduce each line of the table to a line of one of the forms (N_1), (N_2), (N_3).

From each line of form (N_1) let us form an expression $q_i S_j S_k L q_m$; from each line of form (N_2) we form an expression $q_i S_j S_k R q_m$; and from each line of form (N_3) we form an expression $q_i S_j S_k N q_m$ ” (p. 67).

Numbers” (1,2,3,4,5,6 y 7) y letras o “Standar Description” (“A”, “C”, “D”, “L”, “R”, “N”, así como el caracter “;”)²⁰. Con esta máquina se determinaría si los números computables eran enumerables. Eventualmente, debido al problema de la detención (*Halting problem*), Turing concluiría que la hipótesis de la completitud de Hilbert era falsa, lo que se traducía en la imposibilidad de demostrar si el total de las proposiciones eran falsas o verdaderas universalmente²¹.

De acuerdo a esto, la completitud aducida por Hilbert encontraría tres respuestas definitivas que demostraban su falsedad; si bien cada una de ellas establecía que no era posible construir un algoritmo general que valorara si una fórmula del cálculo del primer orden era un teorema, se confirmaría que la solución de Turing y su máquina eran un recurso de demostración válido²²; al respecto Marvin Minsky otorgaría el reconocimiento este procedimiento del siguiente modo:

We define the computable real numbers... as sequences of digits interpreted as decimal fractions. But we add one key restriction. *The digits must be generated sequentially by a Turing machine.* That is, we require that in order that the real number $.a_0a_1a_2\dots$ be a *computable real number*, there must be a Turing machine which starts with a blank tape and prints out a tape of the form:

$$\dots 000Xa_0Xa_1Xa_2X\dots 000^{23}.$$

Como se aprecia, la relevancia de este procedimiento no descansa en la novedad de la conclusión de Turing, sino en los recursos implementados para alcanzar este resultado. La demostración aritmética de Gödel o el cálculo de Church, a pesar de haber argumentado que el teorema de la completitud era falso, se obtenía como logro importante el diseño de una máquina con la cual se demostraba un problema matemático.

A pesar de este avance, el dilema no se hizo esperar: ¿eran las computadoras un recurso que podía sustituir el trabajo de los matemáticos? ¿y a los humanos en

20 TURING, Alan, *The essential Turing*, p. 66.

21 Belen Prado (“Ambivalence in Machine Intelligence”, pp. 66-69) considera que no sólo se tratan de consecuencias para la computación, sino que pueden ampliarse sus resultados a la demostración de que no es posible establecer una determinación para ningún sistema inteligente. Las implicaciones antropológicas, como se ve, son bastante interesantes.

22 Son necesarias varias precisiones sobre esta afirmación, por ello se remite al detalle histórico elaborado por Claudio Gutiérrez (“Kurt Gödel y Alan Turing: una nueva mirada a los límites de lo humano”, *Revista BITS de Ciencia, DCC*, No. 9, 2013, pp. 19-27). La resistencia de los matemáticos de la época, y del propio Gödel, a ciertas generalizaciones son fundamentales en lo que respecta a los fundamentos de las matemáticas y la computación misma. En el siguiente apartado se retoman algunas de estas ideas.

23 MINSKY, Marvin. *Computation. Finite and Infinite machines*, (Nueva Jersey: Prentice Hall, 1967), p. 158, énfasis en el original.

lo general? La cuestión central, al paso de los años, se tradujo en debate importante ya no sólo dentro de las definiciones matemáticas de “función recursiva”²⁴ o dentro del funcionamiento correcto de la propia máquina de Turing y su programación, sino que alcanzó la esfera de la filosofía (antropología y de la mente)²⁵. La cuestión básica apuntaba a si la mecanización de un proceso podía traducirse como la esencia misma del acto intelectual²⁶; si fuese el caso, entonces, la posibilidad para sustituir la inteligencia natural humana por una *AI* no sólo era evidente sino deseable –dados los beneficios que traía consigo resolver los problemas a partir de su mecanización²⁷. Pero la complejidad de estos temas se tratará más adelante, en una nueva sección.

El siguiente apartado presenta los argumentos que ofreció el propio Turing con respecto a la plausibilidad de una máquina capaz de pensar; gran parte de las ideas que propuso estaban focalizadas a satisfacer ciertas posturas sociales que se resistían a aceptar la idea de crear una inteligencia no natural, lo que sirvió para que Turing profundizará en el concepto, proceso metodológico y límites epistemológicos del concepto de *AI*.

3. TURING Y LAS MÁQUINAS INTELIGENTES: ARGUMENTOS DE PLAUSIBILIDAD

A pesar de los avances logrados y de los varios resultados alcanzados por las máquinas que Turing ayudó a diseñar para resolver algunas de las operaciones de comunicación durante la Segunda Guerra Mundial²⁸, la resistencia a una cierta inteligencia cuya naturaleza fuese mecanizada mantuvo por igual el escepticismo como el rechazo entre el mundo matemático. En contramedida, a partir de 1947

24 GUTIÉRREZ, Claudio. “Kurt Gödel y Alan Turing”, pp. 24-27.

25 ROBINSON, William S. “Philosophical Challenges”, (*The Cambridge Companion of Artificial Intelligence*, Edited by Keith Frankish; William M. Ramsey, Cambridge: Cambridge University Press, 2014), pp. 69-72; MINSKY, Marvin. *The Society of Minds*, (New York: Simon & Schuster, 1986), pp. 38-70; BOŁTUĆ, Piotr. “Church-Turing Lovers”, (en: Lin, Patrick; Jenkins, Ryan; Abney, Keith, eds. *Robot Ethics 2.0. From autonomous Cars to Artificial Intelligence*. Oxford: Oxford University Press, 2017), pp. 214-229.

26 ARKOUDAS, Konstantine; BRINGSJORD, Selmer. “Philosophical foundations”, (en: *The Cambridge Companion of Artificial Intelligence*, Edited by Keith Frankish; William M. Ramsey, Cambridge: Cambridge University Press, pp. 36, 63, 2014), pp. 36-39.

27 LIN, Patrick; JENKINS, Ryan; ABNEY, Keith, eds. *Robot Ethics 2.0. From autonomous Cars to Artificial Intelligence*. (Oxford: Oxford University Press, 2017).

28 COPELAND, Jack, *The essential Turing*, pp. 313-335.

Turing comenzó a justificar la posibilidad de las computadoras como objetos físicos capaces de pensar artificialmente²⁹.

Una de las cuestiones centrales –constantemente referidas por el autor– era si la intelección tenía que situarse, necesariamente, al lado de la infalibilidad de las computadoras. El tema no resultaba menor pues si la creación de un sistema inteligente ofrecía mejores recursos de solución a los problemas que por siglos habían ocupado la mente de los investigadores, entonces se erigía la necesidad de reconsiderar las capacidades humanas y, por ende, el nuevo rumbo que podría asumir algunas ciencias como la Matemática. La complejidad estaba en la posibilidad de una computadora para elegir. De acuerdo a su opinión, mientras que una máquina programada no podría frenar su búsqueda ante una solución imposible, un investigador lograría, en cambio, detenerse y elegir una nueva ruta de indagación, lo que reflejaba que la mejor vía era la que ofrecía una intelección natural.

Esta comparación de habilidades entre humanos y computadoras, para Turing, era errónea pues implicaba una inequidad en los términos de valoración de lo que podía entenderse por el concepto de ‘inteligencia’:

It might be argued that there is a fundamental contradiction in the idea of a machine with intelligence. It is certainly true that ‘acting like a machine’ has become synonymous with lack of adaptability. But the reason for this is obvious. Machines in the past have had very little storage, and there has been no question of the machine having any discretion. The argument might however be put into a more aggressive form. It has for instance been shown that with certain logical systems there can be no machine which will distinguish provable formulae of the system from unprovable... if a machine is expected to be infallible, it cannot also be intelligent³⁰.

Como se aprecia, él consideraba que el centro de las críticas se colocaba en la, por entonces, imposibilidad de las máquinas para encontrar vías de solución alternativas dado que la programación inicial impedía una elección más flexible; él mismo advierte que en aquellos casos en los que la incorporación de los recursos permitiese a la computadora considerar algunas estrategias para evitar la circularidad, la designación de un objetivo de solución concentraba las operaciones en un marco operativo delimitado: “The possibility of letting the machine alter its own instructions provides the mechanism for this, but this of course does not get us very far.”³¹.

29 Se recomienda el detalle histórico y sistemático elaborado por Jack Copeland (*The essential Turing*, pp. 217-264) en su edición crítica a las obras de Turing.

30 TURING, Alan, *The essential Turing*, pp. 393-394.

31 TURING, Alan, *The essential Turing*, p. 393.

A pesar de ello, el matemático inglés consideraría que estas restricciones sobre las posibilidades electivas de las computadoras podían encontrar bordes de restricción más amplios en la medida en que el *Hardware* mejorara. La reflexión central se dirigía a encontrar un recurso material que tuviese una memoria lo suficientemente grande como para garantizar que trabajara indefinidamente hasta alcanzar una solución al problema: “it [is] essential in this theoretical arguments that the memory should be infinite”³². Como alternativas para lograr esta idealidad, propone tres rutas de investigación de materiales que hicieran efectiva dicha condición: almacenamiento magnético, patrones de carga en la pantalla de un tubo de rayos catódicos y líneas de retardo acústico como unidades de almacenamiento³³. El esfuerzo de mostrar estas nuevas alternativas para mejorar la memoria, si bien no garantizaban que de ello surgiera una especie de voluntad artificial, se delineaba la idea de que era posible mejorar el marco de recursos electivos derivado de situaciones previas enfrentadas por la máquina y por el programador.

Con la búsqueda de recursos para mejorar la memoria se ampliaba el marco de investigación sobre patrones de aprendizaje, así como el valor del diseño y desarrollo de mecanismos de solución de problemas computacionales. El propio Turing, en 1947, atenuaba este avance para el campo de la computación, pues mesuraba los resultados a los que se podía aspirar con una máquina con memoria casi infinita, ya que consideraba que el papel del analista sería siempre indispensable para mejorar los procesos de la máquina. En contraparte, el beneficio de las computadoras digitales sería la supresión del “sentido común” en ciertas esferas del trabajo humano, pues con la programación de una computada se garantizaba un recurso de operación sistemático: “With a digital computer we can no longer rely on common sense, and the bounds of error must be based on some proved inequalities”³⁴.

Estas y otras indagaciones en la misma línea de reflexión tenían como interés mostrar la plausibilidad del uso dentro de las comunidades matemáticas, por lo que se puede decir que las reflexiones hasta antes de 1940 eran únicamente experimentales. Hacia 1948, en un artículo titulado *Intelligent Machinery*³⁵, propondrá una investigación apegada a las posibilidades reales hasta entonces logradas e iniciará una defensa del desarrollo de una inteligencia no natural. Para tales efectos, comienza por coordinar las objeciones estableciendo que son cinco restricciones contra el diseño de una máquina pensante. Tales son:

32 TURING, Alan, *The essential Turing*, p. 394.

33 TURING, Alan, *The essential Turing*, pp. 380-384.

34 TURING, Alan, *The essential Turing*, p. 391.

35 TURING, Alan, *The essential Turing*, pp. 410-432.

- a) Una renuencia dogmática para aceptar la existencia de un comportamiento inteligente distinto al del ser humano.
- b) La concepción de que religiosamente la creación de una máquina inteligente contravenía a los fundamentos en las que se sostenían tales creencias.
- c) El estado de las máquinas y sus posibilidades de operación alcanzados hasta 1940, lo que, al parecer, demostraba sus límites naturales y necesarios de desarrollo. (Impulsados por una serie de argumentos teológicos, había que garantizar dicho estado a riesgo de intervenir en la forma de la creación deseada por Dios).
- d) La posibilidad limitada por las máquinas de demostrar ciertos teoremas que, por el contrario, los matemáticos podrían resolver ya que podían seguir rutas alternativas de solución reflejándose la capacidad flexible de la mente humana y acentuando la rigidez de las soluciones mecanizadas.
- e) Cualquier tipo de inteligencia 'artificial' no era más que el reflejo de su creador, por ende, la 'inteligencia' correspondía al programador y no podía atribuirse tal cualidad a un objeto material³⁶.

Turing considera que las objeciones (a) y (b), al ser posturas emotivas y carecer de una racionalidad lógica, no requerían respuesta pues carecían de un núcleo argumental; la objeción (c), tampoco requería respuesta pues ya existían máquinas que podían realizar una cantidad de $10^{60.000}$ operaciones sin repetirse, lo que implicaba que, en todo caso, se trataba de una ignorancia sobre el estado alcanzado por la ingeniería computacional hacia 1948. La objeción (d) implicaba un contrasentido, pues de fondo se proponía que una máquina estuviera exenta de errores para demostrarse así su inteligencia siendo que el ser humano había alcanzado un estado de conocimientos importante sólo después de ensayar y corregir sus equivocaciones continuamente a lo largo de la historia. Finalmente, la objeción (e) es refutada diciendo que un alumno no puede ser considerado obra de su maestro una vez el primero logra actuar por cuenta propia o superar al segundo, en el caso de las máquinas, existía evidencia de ambos hechos³⁷.

Una vez Turing resuelve estas objeciones, reflexiona sobre las posibilidades y formas de una *AI*. Él comienza por mostrar sus tipos, los cuales son: Discrete and continuous, Controlling and Active, Logical Computing, Practical Computing, Universal Practical Computing, Paper y, finalmente, Partially random and apparently random machines; la finalidad es mostrar la gran variedad de alternativas de recursos de solución de problemas. Aunado a ello, muestra un cierto tipo

36 TURING, Alan, *The essential Turing*, pp. 410-411; 449-458.

37 TURING, Alan, *The essential Turing*, pp. 412-413.

de máquinas cuya operación dependía de las posibilidades de interrelación de los datos procesados en lugar de la operación como tal. Puesto que el funcionamiento implicaba la vinculación de los datos introducidos (programados), parecía, por ende, que los resultados serían limitados, sin embargo, al centrarse en la interconexión, ello representaría un avance en la metodología de procesamiento de la información. En palabras del propio Turing: “The behavior of a machine with so few units is naturally very trivial. However, machines of this character can behave in a very complicated manner when the number of units is large”³⁸.

Gracias a las posibilidades de interconectar la información, se propondría que una máquina podía alcanzar ciertos estados de procesamiento que el ser humano no era capaz, por lo cual, una computadora inteligente aventajaba en algunas cuestiones a la inteligencia humana; sin embargo, para Turing el objetivo no era la recreación de una máquina humanoide, sino resaltar algunas actividades que él consideraba centrales para el diseño y desarrollo de una *AI*: “...by applying appropriate interference, mimicking education, we should hope to modify the machine until it could be relied on to produce definite reactions to certain commands”³⁹. Para lograr ésto, sitúa las hipótesis de una arquitectura cuyo procesamiento de la información estuviese constituida a partir de un modelo neuronal humano; específicamente consideraba el patrón de flujo cerebral acaecido cuando el individuo experimenta situaciones de dolor y placenteras, pues consideraba que tales experiencias remitían a situaciones modelo para comprender el procesamiento de la información. Específicamente, proponía a la infancia para estudiar el proceso de aprendizaje durante el crecimiento⁴⁰.

El autor propuso desarrollar una línea de investigación para construir una máquina capaz de operar con patrones aleatorios, basándose en el modo en que el cerebro humano procesa cierta información; con ello propuso diseñar un sistema de alimentación *input-output* y un sistema de procesamiento semejante a la estructura neuronal del modo en que operan las redes neuronales. Lo relevante filosóficamente de esta hipótesis era la idea de la existencia de una simetría conceptual entre el diseño de una máquina (programación) y el crecimiento (aprendizaje) de un infante, pues de ese modo se proponía la unidad entre el mundo teórico, el de la percepción y el natural. Turing, en una delimitación a sus alcances, resume así sus reflexiones:

The analogy with the human brain is used as a guiding principle. It is pointed out that the potentialities of the human intelligence can only be realized if suitable

38 TURING, Alan, *The essential Turing*, p. 417.

39 TURING, Alan, *The essential Turing*, p. 422.

40 TURING, Alan, *The essential Turing*, pp. 422-428.

education is provided. The investigation mainly centers round an analogous teaching process applied to machines. The idea of an unorganized machine is defined, and it is suggested that the infant human cortex is of this nature... In one case the education process is carried through until the organization is similar to that of an [Automatic Computing Engine]⁴¹.

En general, era claro que para el matemático la analogía humano-computadora debía aparejarse tanto como hipótesis de experimentación e investigación, como modelo de desarrollo para el desarrollo de una cierta forma artificial de inteligencia deseable para la creación de una nueva generación de computadoras.

Estas reflexiones se verán profundizadas en otro celebre texto de 1950 titulado *Computing Machinery and Intelligence* en donde sugiere el famoso "Imitation game"⁴²; la tesis central del artículo plantea el siguiente experimento mental: supongamos que tres personas, (A) una mujer, (B) un hombre y un (C) participante (sexo indistinto) que no sabe si A o B es la mujer. Como parte de las condiciones del juego se encuentran aislados unos de los otros y se encuentran comunicados por tarjetas mecanografiadas que pueden enviarse entre sí. El objetivo del juego es determinar, por parte de C, cuál de los participantes es la mujer (A o B). La cuestión se vuelve más compleja cuando Turing propone la siguiente variación:

'What will happen when a machine takes the part of A in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, 'Can machines think?'⁴³.

Si bien la recepción de la experimentación cuestionaba la identidad entre la definición de inteligencia, el procesamiento del lenguaje y la representación conceptual lograda de dicha vinculación⁴⁴, la propuesta de Turing se seguía manteniendo en la exploración de los procesos de aprendizaje infantil y las posibilidades de modelar procesos cognitivos a partir de dicho patrón de desarrollo evolutivo. De acuerdo con él, la evidencia psicopedagógica demostraba que la inteligencia infantil dependía de los estímulos que el medio circundante le ofrecían (input) para producir estados conductuales específicos (output) de tal manera que el procesamiento de la información (aquello que se socioculturalmente se reconoce como 'Inteligencia'), estaba directamente vinculado con la capacidad del agente

41 TURING, Alan, *The essential Turing*, pp. 431-432.

42 TURING, Alan, *The essential Turing*, pp. 441-464.

43 TURING, Alan, *The essential Turing*, p. 441.

44 DREYFUS, Hubert, *What Computers can't do. A critique of Artificial Reason*, (New York: Harper & Row, 1973), en adelante: *Computers*; SEARLE, John. "Minds, Brains, and Programs", (*The Behavioral and Brain Sciences*, Vol. 3 (1980), pp. 417-457), en adelante *Minds*; ARKOUDAS, Konstantine; BRINGSJORD, Selmer. "Philosophical foundations", pp. 46-52.

para procesar una cierta información recibida y, eventualmente, responder en consistencia con dicha proposición. En ese contexto, para Turing⁴⁵, la conclusión era evidente: si un ser humano determina que las respuestas dadas por una computadora digital provienen de un ‘ser humano’, entonces había que asumir que la máquina era inteligente.

La conclusión, como se aprecia, asumía al menos dos referentes epistémicos centrales para construirse, los cuales podrían formularse de la forma siguiente:

- a) La inteligencia es el producto de una formulación lingüística consistente con ciertos parámetros culturalmente constituidos.
- b) El desarrollo cerebral, sobre todo durante el proceso de maduración infantil, muestra una estructura base del intelecto humano, el cual, por ende, sirve como base para la arquitectura del software.

Para el autor era fundamental que el participante humano determinara que una computadora digital era específicamente un ‘ser humano’ (asumiendo tal respuesta a partir de las respuestas dadas por la máquina), pues con ello se implicaba que un agente epistémico reconocía un conjunto de patrones conceptuales aprendidos culturalmente (y atribuidos convencionalmente a la inteligencia humana) en los textos intercambiados. Para alcanzar esta imitación, Turing proponía asumir a las redes neuronales como un criterio de desarrollo, pues nuclearmente el proceso de aprendizaje moldeaba el cerebro del niño de acuerdo a ciertos modelos culturales para que actuara bajo patrones conductuales reconocidos como inteligentes. Por ello, él mismo concluiría lo siguiente: “Intelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation, but a rather slight one, which does not give rise to random behaviour, or to pointless repetitive loops”⁴⁶; a partir de esto, será fundamental explorar los procesos y patrones de aprendizaje para mejorar las respuestas inteligentes esperables de una computadora, pues de los procesos de programación se lograría una mejor respuesta por parte de la máquina.

Como se aprecia, para el autor la idea de una computadora inteligente no constituía una simple analogía semántica o algún tipo de metáfora para referirse a una nueva área de la computación, en cambio implicaba el esfuerzo real para examinar los modelos de investigación relacionados con esta área de conocimiento: “My contention is that machines can be constructed which will simulate the behaviour of the human mind very closely”⁴⁷. Es importante remarcar que la finalidad última de Turing no era otra que demostrar lo siguiente: “machines will eventually

45 TURING, Alan, *The essential Turing*, pp. 448-449.

46 TURING, Alan, *The essential Turing*, p. 463.

47 TURING, Alan, “Intelligent Machinery. A Heretical Theory”, *The essential Turing*, p. 472.

compete with men in all purely intellectual fields”⁴⁸. En ese sentido, no es extraño que hubiese impulsado una reflexión sobre las condiciones de posibilidad a través de las cuales las máquinas fuesen capaces de pensamiento y cognición.

Hasta aquí se han mostrado los argumentos aducidos por Turing con respecto a la posibilidad de crear una computadora inteligente. Se ha visto que parte de la apología se proponía justificar los fundamentos y estructura de una máquina pensante procurando mostrar la factibilidad antropológica y ética de crear un objeto material capaz de manifestar tanto cognición como actividad mental. En el siguiente apartado se muestran algunos de los debates y lecturas desprendidos de las propuestas de Turing. La finalidad de esta selección es mostrar una doble cuestión: primero, los cuestionamientos por parte de matemáticos (inicialmente) y filósofos (posteriormente) a la factibilidad de que una máquina resolviese problemas o que fuese capaz de pensamiento. La segunda cuestión apuntará a mostrar cómo el proyecto de construir una computadora inteligente gradualmente se orientó a la configuración del término *AI*.

4. LA INTELIGENCIA ARTIFICIAL SEGÚN TURING: LIMITACIONES Y RELECTURAS

La presente sección se subdividirá en dos sub-apartados para mostrar algunas limitaciones a las posibilidades de una computadora inteligente: revisiones de matemáticos y relecturas filosóficas. A través de la presentación de estas reflexiones que se harían sobre las propuestas de Turing, se valoraran los aportes definitivos, las limitaciones y los fundamentos que servirían para una posterior definición del concepto de *AI*.

4.1. EL TRABAJO Y LA FLEXIBILIDAD DEL PENSAMIENTO MATEMÁTICO

La propuesta de una máquina que determinara si la hipótesis de Hilbert era posible, aunque encontró el respaldo y validación de Alonzo Church en su momento, ello no implicaba que la comunidad estuviese de acuerdo en la totalidad de los argumentos y, sobre todo, en los procesos que Turing proponía para demostrar la falsedad de la proposición⁴⁹. Casi de inmediato Emil Post publicaría una crítica al estudio sobre los números computables; algunas de las restricciones pro-

48 TURING, Alan, *The essential Turing*, p. 463.

49 Cfr. COPELAND, Jack, en: *The essential Turing*, pp. 52-53; 91-93.

puestas por el autor serían: (1) algunos de los comandos propuestos no presentan los resultados determinados por Turing, ya sea porque sus cálculos no fueron precisos o porque habían algunos errores menores de notación; (2) se establecían algunos procedimientos iniciales para el movimiento (la máquina únicamente se puede mover a la izquierda al iniciar, nunca borra el símbolo dado inicialmente, ni imprime alguno cuya definición estuviese dada; Post se pregunta si esta decisión arbitraria incidía en la demostración de la insolubilidad de la hipótesis de Hilbert. (3) La tercera crítica, remite, nuevamente, a la selección sin criterios definidos de los símbolos cuyo cálculo implicaba que la máquina de Turing operara de un distinto al definido inicialmente⁵⁰.

En la misma línea, Donald Davies, presentaría un informe detallando dichos errores en la redacción original de Turing, al tiempo que insistía que éstas provocarían confusiones entre los lectores: “I realized that, even though the feasibility of the universal computing machine was not in doubt, the mistakes in Turing’s exposition could puzzle future readers and plague anyone who tried to verify Turing’s design by implementing his universal machine in practice”⁵¹. La respuesta de Turing, en una carta a Alonzo Church, explicaría que estaba de acuerdo en la existencia de dichos errores, sin embargo, reafirmaba que las fallas señaladas a su trabajo no alteraban el funcionamiento general propuesto.

La relevancia de estas precisiones remitía a la viabilidad de confirmar el teorema de la incompletitud de Gödel a través de medios computacionales. De acuerdo con Jack Copeland⁵², editor de las obras de Turing, algunos matemáticos se opusieron al surgimiento de una máquina como vía de demostración de teoremas, el argumento central era que se trataba de una tarea humana que no podría asumir a una computadora⁵³. De acuerdo con Claudio Gutiérrez⁵⁴, la tensión central estaba tanto en las aplicaciones que proponía, como en los alcances que la máquina de Turing ofrecía para la demostración de teoremas de primer orden. Si bien la aportación de una máquina que operara de acuerdo a las especificaciones del matemático inglés no aportaban un conocimiento distinto al desarrollado por Gödel en 1931, sí, al menos, ofrecía un criterio nuevo de interpretación de los mismos resultados presentados por el matemático alemán: la demostración de que la elección de los axiomas puede seguir un patrón intuitivo y, por tanto, su justificación proviene de la efectividad de la operación de una

50 POST, Emil, en: *The essential Turing*, pp. 98-99.

51 DAVIES, Donald, *The essential Turing*, p. 103.

52 COPELAND, Jack, *The essential Turing*, p. 53.

53 PRADO, Belen, “Ambivalence in Machine Intelligence”, pp. 61-64.

54 GUTIÉRREZ, Claudio, “Kurt Gödel y Alan Turing”, pp. 24-27.

máquina programada bajo un esquema también intuitivo. Precisamente esta es una de las objeciones de Gödel:

The [lack of statements] necessarily presuppose a kind of Platonism, which cannot satisfy any critical mind and which does not even produce the conviction that they are consistent⁵⁵.

Al hablar de “platonismo” Gödel apuntaba a que un sistema de la forma S justificaba un conjunto de teoremas $s_1, s_2, s_3 \dots s_n$ a partir los argumentos derivados de S , lo que haría que el sistema operara bajo un “idealismo formal” indemostrable⁵⁶. De acuerdo con Turing, las computadoras inteligentes, si bien proponían un sistema de programación para operar de acuerdo a criterios propuestos por el propio programador (sin necesariamente ofrecer una justificación inicial de dicha elección), ello no implicaba que los resultados carecieran de fundamentos demostrables, pero en lugar de atribuir argumentos en favor de su plausibilidad, el propio funcionamiento mostraría si el código de programación era correcto.

En resumen, se aprecia que las objeciones matemáticas se dirigían a cuestionar las posibilidades de una máquina que fuese capaz de inteligencia, pues se aducía que un proceso mecánico se ajustaba la rigidez de la programación predefinida y que, además, parecía no justificar los principios de los que partía. Turing, al respecto, objetará dos cuestiones: primera, que las posibilidades de una computadora inteligente no deben valorarse a partir de la ‘infalibilidad’, pues esta no es una característica de la inteligencia natural⁵⁷; en ese sentido, las posibilidades correctivas (ensayo y error) permitían la reestructuración de patrones de modelación cada vez más precisos. Segunda, aun cuando la teoría de la incompletitud apuntaba la necesidad de incorporar sistemas o funciones alternativos (S') para evaluar la factibilidad de un sistema dado (S), para el caso de la computación había que, primero, delimitar el campo semántico de posibles problemas que podían ser resueltos por una computadora; estratégicamente Turing proponía que la delimitación de los problemas cuya solución aceptaba métodos computacionales, no sólo serviría para definir el propio campo epistemológico de esta área

55 GÖDEL, Kurt, “On undecidable sentences” (en: Solomon Feferman, John Dawson, Warren Goldfarb, Charles Parsons, Robert Solovay, eds., *Kurt Gödel. Collected Works*, Vol. III, Oxford: Oxford University press), p. 35. En adelante: *Kurt Gödel*, III.

56 *Kurt Gödel*, III, “The present situation in the foundations of mathematics”, p. 50.

57 El propio Turing explica que sólo se puede saber lo que una computadora realizará hasta que finalice su procedimiento, pues si el programador fuese capaz de conocer el resultado, el acto mismo de programar carecería de sentido: “Sometimes a computing machine does do something rather weird that we hadn’t expected. In principle one could have predicted it, but in practice it’s usually too much trouble. Obviously if one were to predict everything a computer was going to do one might just as well do without it”, *The essential Turing*, p. 500.

de conocimiento, sino para garantizar la universalidad del método empleado así como el resultado alcanzado.

Con esto Turing ofrecería una justificación al término maquina o computadora “inteligente”, pues, así como la práctica de la matemática no estaba exenta de errores y, sin embargo, a través del gradual avance de las investigaciones se habían encontrado soluciones a los problemas que por años pudieron parecer irresolubles, de modo análogo, no se podía esperar que un campo emergente como el de la computación ofreciera respuestas definitivas a todas las cuestiones sin ofrecerle un periodo de tiempo suficiente para que los resultados se clarificaran y, con ello, pudiera definirse con precisión la agenda de problemas propios de una actividad vinculada al uso e investigación de las computadoras inteligentes. Al respecto, el propio Turing en 1951 plantearía la siguiente idea:

It might be argued that there is a fundamental contradiction in the idea of a machine with intelligence... It is certainly true that ‘acting like a machine’ has become synonymous with lack of adaptability... Against it I would say that fair play must be given to the machine. Instead of it sometimes giving no answer we could arrange that it gives occasional wrong answers... in other words then, if a machine is expected to be infallible it cannot also be intelligent... No man adds very much to the body of knowledge, why should we expect more of a machine?⁵⁸

Precisiones como esta, a pesar de su simpleza, aclaraban tanto el rumbo hacia el que debía orientarse una agenda de problemas relacionados con el estudio de la “inteligencia” de las computadoras, como el modelo del que debía partir para alcanzar resultados sólidos y definitivos. Esta apología, no podría resultar más profética, pues como se mencionó en la introducción apenas dos años después de su muerte en 1954, en Darmouth, un grupo de investigadores encabezados por McCarthy definirían el campo de trabajo de una *AI* reconociendo el funcionamiento del cerebro humano como modelo de procesamiento de información como referencia, pero también explicando que se trataba de objetivos propios e independientes de otras ciencias, pues a pesar de que se nutría de, por ejemplo, la Matemática o la Neurociencias, tenía una ruta de investigación propia y con tareas a desarrollar; dicha independencia, debe mucho a las defensas y proposiciones que realizaría Turing a lo largo de su actividad profesional. Pasamos ahora a revisar las objeciones filosóficas a las que fueron sometidas las propuestas del autor.

58 TURING, Alan, *The essential Turing*, p. 394.

4.2. SENTIDO Y FORMA DE UNA INTELIGENCIA ARTIFICIAL ¿PUEDEN LAS MÁQUINAS ENTENDER?

Si Turing había resuelto que la circunscripción de los temas y problemas posibles de una disciplina que investigara la “inteligencia” de las computadoras, debía ajustarse a una agenda propia de problemas cuyo lenguaje, metodología y objetivos se entendieran dentro de sus propios términos, para los filósofos el problema no se zanjaría únicamente por establecer límites epistemológicos y metodológicos entre los investigadores en el área computacional y el resto de saberes. Aunque son varios los dilemas que se pueden establecer⁵⁹, uno de los más importantes debates remitió al sentido y forma del significado de una forma de inteligencia no natural. Es claro que Turing, entre 1948 y 1950 defendió la idea de una cierta capacidad intelectual que podía ser reproducida en objetos materiales una vez se asumiera el modelo neuronal de desarrollo en la etapa infantil. De esta comparación análoga, a los filósofos les interesaría principalmente las implicaciones cognitivas: ¿puede una estructura diseñada a partir de un modelo neuronal humano entender?

Hubert Dreyfus y John Searle abrirían el debate al reconocer que en un cierto nivel la comprensión blanda y fuerte de la *AI* (*weak and strong AI*)⁶⁰ ayudarían a clasificar el grado de racionalidad viable de modelación neuronal para la organización de un programa computacional. La oposición central era suponer que, necesariamente, el cerebro humano operaba como una computadora y, por ende, había que asumir a la inteligencia como un cierto tipo de programa avanzado que habilitaba el entendimiento humano. Hay que recordar que Turing consideraba lo siguiente: “our digital computer, suitably programmed, will behave like a brain”⁶¹. A partir de ello, Dreyfus, años más tarde, sostendría:

a neural net can be simulated using a program, but such a program is in no sense a heuristic program. Thus the mere fact that the brain might be a digital computer is in no way ground for optimism as to the success of artificial intelligence as defined by Simon or Minsky⁶².

59 Para una síntesis de estos problemas filosóficos se recomiendan los siguientes trabajos: ARKOUDAS, Konstantine, pp. 34-63; BRINGSJORD, Selmer. “Philosophical foundations”; ROBINSON, William S. “Philosophical Challenges”, pp. 64-88; Prado, Belén, “Ambivalence in Machine Intelligence”, pp. 57-61.

60 El término es propuesto por Searle (*Minds*, p. 417), pero se aprecia que Dreyfus (*Computers*, pp. 71-72) acepta a un cierto nivel la modelación neuronal como referente para el desarrollo de ciertos sistemas de procesamiento de información.

61 TURING, Alan, *The Essential Turing*, p. 483.

62 DREYFUS, Hubert, *Computers*, p. 71.

Por su parte, para Searle:

according to strong *AI*, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to *understand* and have other cognitive states⁶³.

Como se aprecia, la crítica central no se remitía a la posibilidad de diseñar y desarrollar a las computadoras a partir de un modelo neuronal, sino, justamente, invertir los términos: que el cerebro fuese entendido a partir de un descubrimiento reciente generaba en ambos autores una primera oposición; ¿Qué criterios justificaban el definir la intelección natural a partir de una cierta intelección artificial? Pero la cuestión, más allá del anacronismo, recaía directamente en asumir que las computadoras eran objetos materiales que, debido al procesamiento de la información, debían reconocerse como (i) agentes epistémicos, (ii) capaces de entendimiento. Específicamente Dreyfus se opondrá a esta cuestión al señalar enfáticamente que “the evidence is against the possibility of using digital computers to produce intelligence”⁶⁴. Searle hará eco de la objeción al sostener que “the claim that the appropriately programmed computer literally has cognitive states and that the programs thereby explain human cognition”⁶⁵. La resistencia filosófica se oponía a la analogía entre un cierto modo de procesar la información cuyo referente de diseño consideraba la organización neuronal.

Estas críticas de Dreyfus y Searle tienen como base las posturas de Turing, para quien la estructuración de un programa que reproducía el modelo neuronal para procesar la información debía, por tanto, esperar resultados similares a los alcanzados por el cerebro humano. Es cierto que él aceptaba que el estado de la *AI* se encontraba en un momento incipiente, pero no por ello se debía desacreditar el rápido avance hacia el que se dirigiría la evolución de esta área de estudios. De acuerdo a ello, entonces, existía un contexto de racionalidad a través del cual se podía atribuir que las computadoras, independientemente de lo que culturalmente se definiera como procesamiento de información, no contravenía a que una *AI* ofrecía resultados similares a los esperados para un agente epistémico: “The idea... is that the machine has to try and pretend to be a man, by answering questions put to it, and it will only pass if the pretence is reasonably convincing”⁶⁶.

63 SEARLE, John, *Minds*, p. 417.

64 DREYFUS, Hubert, *Computers*, p. 74.

65 SEARLE, John, *Minds*, p. 417.

66 TURING, Alan, “Can Automatic Calculating Machines Be Said to Think?”, *The essential Turing*, p. 495. El objetivo del diálogo en el que se enmarca el presente texto citado es, precisamente, definir el grado de comparación entre la forma de la inteligencia humana y la arquitectura de programación inteligente de una computadora. Además, es aquí donde Turing proyecta que para que

Una de las cuestiones troncales de esta equivocidad provenía de las dificultades para definir el concepto de “aprendizaje”. La introducción de este nuevo término se justificaba porque si el acto intelectual se configuraba a través de un proceso de aprendizaje a partir de la relación “*Inteligencia si sólo si Aprendizaje*” ($I \leftrightarrow A$), entonces, para Turing, era posible revertir el estado bicondicional de tal modo que la relación se mantenía al sostener que $A \leftrightarrow I$. Al invertir los términos, era aceptable aducir que se habilitaba una ruta de investigación a partir de las condiciones generales que permitían al ser humano aprender. Considerando esta ruta, Turing proponía que dada una memoria lo suficientemente grande como para almacenar una buena cantidad de datos y un diseño *Hardware* que simulara una red neuronal, restaba por desarrollar la arquitectura de programación que reprodujera operaciones consideradas como inteligentes por un ser humano.

Dreyfus⁶⁷ se opondrá a estas ideas al sostener que se suele asumir que el acto reflexivo, así como el propio acto cognitivo de ordenamiento de la información es considerado evidente por parte de buena parte de los investigadores de la AI, por lo que parecería que no hay razones para dudar de una conducta ‘inteligente’. Para el autor, en cambio, dichas actividades, filosóficamente hablando, requieren una amplia justificación para garantizar tanto origen, desarrollo como interrelación de la información con la cual expresamos buena parte de nuestras emociones estados cognitivos básicos (miedo, creencias, seguridad, certeza, etc.). De acuerdo a ello, lo inteligente no es evidente y, por ende, la relación analógica entre el actuar de una computadora y el del ser humano no implicaba que una acción fuese ‘inteligente’.

La respuesta a estas objeciones, provendrían de, precisamente, los impulsores del término AI en Darmouth en 1956. De acuerdo con Marvin Minsky –un “entusiasta” de la AI⁶⁸, si bien existen algunos estatutos básicos atribuibles a esta facultad humana (resolución de problemas complejos, capacidad de adaptación, autoconciencia, entre otros), con el progresivo descubrimiento de resultados desde

las computadoras puedan superar el denominado *Test de Turing* deberán pasar, al menos, 100 años (para 2050 aproximadamente).

67 DREYFUS, Hubert, *Computers*, pp. 75-81.

68 Dreyfus (*Computers*) califica así, entre otros a Minsky, de tal modo que las objeciones están dirigidas tanto a la AI y a los investigadores entusiastas de esta área de estudio: “In 1957 Simon predicted that within ten years psychological theories would take the form of computer programs, and he set out to fulfill this prediction by writing a series of programs which were meant to simulate human cognition by simulating the conscious and unconscious steps a person goes through to arrive at a specific cognitive performance. And we have seen that despite the general inadequacy of such programs, admitted even by enthusiasts such as Minsky, all those involved in the general area of artificial intelligence (Minsky included) share the assumption that human beings, when behaving intelligently, are following heuristic rules similar to those which would be necessary to enable a digital computer to produce the same behavior”, p. 76.

áreas interdisciplinarias demostraba que las concepciones de lo ‘inteligente’ podía reajustarse a nuevos marcos de comprensión, lo que implicaría que no todo acto intelectual era reducible a un estatuto natural de nuestra capacidad. En todo caso, se trata tanto de una actividad natural propia del ser humano y de una concepción que reconoce la interconectividad de funciones como ‘acción inteligente’⁶⁹.

Además, a diferencia de Dreyfus, Minsky sostiene que un sistema inteligente no sólo remite a un sistema cuya operación es capaz de resolver problemas que ninguna “mente humana” puede, eso sólo implicaría una región desconocida por explorar a fin de encontrar una respuesta. En contramedida, para Minsky un sistema “experto” es cualquier actividad cuya programación implica la interrelación de información que habilita acciones consideradas “de sentido común (de acción regular)”; sin embargo, la facilidad de una acción, únicamente demuestra la gran capacidad neuronal del ser humano para resolver situaciones complejas, a pesar calificarse como una resolución simple. En ese sentido, el que se pueda programar una computadora para realizar acciones “de sentido común” refleja un gran avance en el campo de la *AI*:

To be considered an ‘expert,’ one needs a large amount of knowledge of only a relatively few varieties. In contrast, an ordinary person’s ‘common sense’ involves a much larger variety of different types of knowledge-and this requires more complicated management systems⁷⁰.

Otros argumentos en favor de situar una *AI* como un campo que había logrado una independencia luego de las propuestas de Turing, fue McCarthy quien –en su respuesta a las críticas de Dreyfus⁷¹– consideraba que una aportación que se podía esperar de las computadoras era una nueva clase de “intuicionismo” para la fundamentación del conocimiento, particularmente de la lógica. En lugar de entender este proceso cognitivo como un acto creativo sobre el cual se pretendía sostener la programación de las computadoras se debía asumir como un proceso sistemático que implicaba un modo de proceder organizado, pero del cual únicamente se establecían la lógica de procedimiento general, para luego esperar el resultado de una computadora con la cual se simplificaban las operaciones, pero no así los principios que debían procesarse por parte del ordenador: “Common sense knowledge and reasoning is at the core of *AI*, because a human or an intelligent machine always starts from a situation in which the information available to it has a common sense character”⁷².

69 MINSKY, Marvin, *The Society of Mind*, (Londres Pan Books, 1988), p. 71.

70 MINSKY, Marvin, *The Society*, p. 72. Énfasis en el original.

71 Se refiere particularmente a los planteamientos realizados en *Computers*.

72 McCARTHY, John, *Defending AI Research*, p. 89.

Con todo esto, se aprecia una cuestión fundamental que había sido defendida por Turing y que luego sería objeto de debate en el seno filosófico y por parte de quienes propusieron formalmente el término *AI*; el dilema central indagaba si el concepto de Inteligencia, de acuerdo a la concepción de los filósofos, necesariamente se ligaba a las acciones humanas vinculadas con el ejercicio del conocimiento, o si, en cambio, remitía a la capacidad para procesar información asumiendo como modelo de procesamiento a la actividad cerebral. Siendo esta una cuestión que se ha reflexionado aún después de la muerte de Turing, se puede pensar por qué el matemático inglés encontró resistencias cuando propuso la idea de una máquina capaz de entendimiento. Pero aun cuando éste último no hubiese encontrado una respuesta definitiva, sin duda sus aportaciones resultaron cruciales para el desarrollo del término, así como para la delimitación de las tareas propias de este campo.

En resumen, el concepto de *inteligencia* y *entendimiento* que Turing consideraba podían desarrollar una máquina, remitían a las posibilidades de interrelación de datos e información que se podía habilitar para que una computadora operara. Si bien la principal crítica de Dreyfus y Searle se enfocan a cuestionar que no todo sistema cuya arquitectura tome como modelo el cerebro humano operará inteligentemente, ni, por ende, entenderá la información que procesa, para Turing (y posteriormente McCarthy y Minsky) la conducta inteligente remite a la posibilidad de interrelacionar la información a través de un modelo neuronal de interconexión. En ese sentido, la cuestión artificial de este modelo de comportamiento, no se propone alcanzar conductas humanas por parte de una computadora, sino programar acciones cuyo procesamiento informacional (por simple que este pueda ser) sea análogo al modo en que es procesado por un ser humano.

CONCLUSIONES

Aunque no existe un concepto de *Artificial Intelligence* (AI) propuesto por Turing, se aproxima a él a través de su propuesta de una *máquina pensante*, y la cual, a su vez, fue determinante para el surgimiento del término en 1956, en Darmouth en el encuentro entre McCarthy, Minsky y otros. De acuerdo con lo desarrollado hasta aquí, tres cuestiones fueron las aportaciones de Turing para la eventual definición del concepto: (1) el concepto de *Inteligencia* no remite, necesariamente, a la esfera humana pues esta capacidad refiere a la posibilidad de resolver problemas a partir de un sistema de principios definidos inicialmente y a partir de los cuales se operarán y definirán acciones posteriores. (2) Una *máquina* es capaz de procesar información a partir de la delimitación de funciones y

operaciones generales, lo cual cumple con la definición de inteligencia. (3) El progreso de las investigaciones eventualmente cristalizaría en un campo más definido para el desarrollo de los problemas relacionados con la intelección (resolución de problemas) y la computación (máquinas que procesarían y calcularían información) para con ello simplificar procesos cuyas aplicaciones podrían favorecer a distintos campos de saberes.

El proyecto de Turing, a pesar de ofrecer soluciones a diversos campos de problemas en la matemática y la criptografía, por citar ejemplos, encontró resistencias a su concepción de un campo vinculado con las máquinas inteligentes. Dos, fueron las principales resistencias: por parte de los *matemáticos*, las objeciones cuestionaban los procedimientos lógicos y metodológicos de entre las cuales una computadora elegía para proceder a completar una demostración; la resistencia central estaba, por tanto, en la falta de flexibilidad inherente a la línea de programación. Los *filósofos*, por su parte, se oponían al concepto “inteligencia” pues implicaba la capacidad volitiva y cognitiva atribuible a un objeto material que, por definición, carecía de capacidad para actuar bajo parámetros humanos. Sin embargo, la aportación central de Turing remite al desarrollo de un sistema de procesamiento de la información a través de la interconectividad de datos con los cuales se habilite una toma de decisiones cuyo resultado favorezca la resolución de un problema que requiere la mecanización de los procesos. Sin embargo, la paradoja central de este modo de proceder es que a pesar de que existe una programación predefinida, aún con ello no se garantiza que se puedan predecir los resultados, lo cual implica que el procesamiento de la información con *AI* cumple con las condiciones básicas esperadas para una inteligencia humana: la toma de decisiones sigue siendo indeterminable e imposible de definir a pesar de que opera con datos e instrucciones conocidos; una cuestión que sin justificar que las computadoras son capaces de pensar, sí, al menos, habilita la posibilidad de responder afirmativamente a la pregunta de Turing: “Can Machines think?” y que años más tarde permitiría el surgimiento de un campo denominado *Artificial Intelligence*.

REFERENCIAS BIBLIOGRÁFICAS

- ARKOUDAS, Konstantine y BRINGSJORD, Selmer, “Philosophical foundations”. En: FRANKISH, Keith; RAMSEY, William M., eds., *The Cambridge Companion of Artificial Intelligence*, Cambridge University Press, 2014, 34-63.
- BOŁTUĆ, Piotr, “Church-Turing Lovers”. En: LIN, Patrick; JENKINS, Ryan; ABNEY, Keith, eds. *Robot Ethics 2.0. From autonomous Cars to Artificial Intelligence*. Oxford: Oxford University Press, 2017, 214-229.
- CHURCH, Alonzo, “A Note on *Entscheidungsproblem*”, *The Journal of Symbolic Logic*, Vol. 1, No.1, 1936, 40-41.

- , *Introduction to Mathematical Logic. Vol. I.* New Jersey: Princeton University Press, 1956.
- DAVIES, Donald W., “Corrections to Turing’s Universal Computing Machine”. En: TURING, Alan. *The essential Turing. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secret of Enigma.* Edited by Jack Copeland. Oxford: Clanderon Press/Oxford University Press, 2004, 103-124.
- DREYFUS, Hubert, *What Computers can’t do. A critique of Artificial Reason*, (New York: Harper & Row, 1973).
- GARCÍA, Javier, *La máquina de Turing (explicada)*. Conferencia 2 de diciembre de 2016, en: <https://www.youtube.com/watch?v=NS-NQ5mCSs8>
- GÖDEL, Kurt, “On undecidable sentences”. En: FEFERMAN, Solomon; DAWSON, John; GOLDFARB, Warren; PARSONS, Charles, SOLOVAY, Robert, eds., *Kurt Gödel. Collected Works*, Vol. III, Oxford: Oxford University press, 1995, 30-35.
- , “On formally undecidable propositions of *Principia mathematica* and related systems I”. En: GÖDEL, Kurt, *Collected Works, Vol. I (Publications 1929-1936)*. Edited by Solomon Feferman. Oxford: Oxford University Press, 1986, 144-195.
- GUTIÉRREZ, Claudio, “Kurt Gödel y Alan Turing: una nueva mirada a los límites de lo humano”. *Revista BITS de Ciencia, DCC*, No. 9, 2013, 19-27. Disponible en: <https://users.dcc.uchile.cl/~cgutierr/otros/godel-turing-bits.pdf>
- HILBERT, David, *Mathematical Problems*. ICM París, 1990.
- , “On the Foundations of mathematics”. En: HEIJENOORT, Jean van, ed. *From Frege to Gödel. A Source Book in Mathematical Logic, 1879-1931*. Cambridge: Harvard University Press, 1967, 464-479.
- LIN, Patrick; JENKINS, Ryan y ABNEY, Keith (eds.), *Robot Ethics 2.0. From autonomous Cars to Artificial Intelligence*. Oxford: Oxford University Press, 2017.
- McCARTHY, John; MINSKY, Marvin; ROCHESTER, Nathaniel y SHANNON, Claude, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, p. 1. Disponible en línea: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf> (última consulta 2 de agosto de 2021).
- McCARTHY, John, *Defending AI Research. A collection of Essays and Reviews*. Stanford: CSLI Publications, 1996.
- MINSKY, Marvin, *Computation. Finite and Infinite machines*. Nueva Jersey: Prentice Hall, 1967.
- , *The Society of Minds*. New York: Simon & Schuster, 1986.
- POST, Emil, “On computable numbers, with an Application to the *Entscheidungsproblem*. A critique”. En: TURING, Alan. *The essential Turing*, 97-101.
- PRADO, Belen, “Ambivalence in Machine Intelligence: The epistemological Roots of the Turing Machine”, *Signos Filosóficos*, Vol. 23, No. 45, 2021, 54-73.
- ROBINSON, William S., “Philosophical Challenges”, *The Cambridge Companion of Artificial Intelligence*, Edited by Keith Frankish; William M. Ramsey, Cambridge: Cambridge University Press, 2014, 64-88.
- SEARLE, John, “Minds, Brains, and Programs”, *The Behavioral and Brain Sciences*, Vol. 3, 1980, 417-457.

TURING, Alan, *The essential Turing. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secret of Enigma*. Edited by Jack Copeland. Oxford: Clarendon Press/Oxford University Press, 2004.