Sophia, Colección de Filosofía de la Educación
ISSN: 1390-3861
ISSN: 1390-8626
revista-sophia@ups.edu.ec
Universidad Politécnica Salesiana
Ecuador

González Fernández, Rodrigo Alfonso
La factibilidad del niño-máquina de Turing
Sophia, Colección de Filosofía de la Educación,
núm. 39, 2025, Julio-Diciembre 2026, pp. 115-139
Universidad Politécnica Salesiana
Cuenca, Ecuador

DOI: https://doi.org/10.17163/soph.n39.2025.03

https://doi.org/10.17163/soph.n39.2025.03

La factibilidad del niño-máquina de Turing

The Feasibility of Turing's Child-Machine

Rodrigo Alfonso González Fernández[*]
Universidad de Chile, Santiago de Chile, Chile
rodgonza@uchile.cl
http://orcid.org/0000-0001-9693-0541

Forma sugerida de citar: González Fernández, Rodrigo Alfonso. (2025). La factibilidad del niño-máquina de Turing. *Sophia, Colección de Filosofía de la Educación*, (39), pp. 115-139.

Resumen

La máquina de Turing (MT) y el controvertido "juego de la imitación" son los aportes más reconocidos de Alan Turing en la filosofía de la inteligencia artificial (IA). El proyecto del niño-máquina, directamente ligado al aprendizaje de máquinas programadas o computadores digitales, no es tan reconocido, aunque no es menos importante. Según dicho proyecto, una máquina programada debe aprender como un niño, si ha de convertirse en una "mente adulta pensante", esto es, que juzgue, que entienda, que distinga. En este artículo se muestra que ese desiderátum de Turing no es realizable mediante algoritmos. Mientras que en la primera sección se introduce al problema, en la segunda se da un breve recuento histórico de los algoritmos, las máquinas de Turing y de su relación. En la tercera sección se aborda el concepto *machine intelligence* y el proyecto del niño-máquina. En la cuarta sección se muestra que una forma de entendimiento ("la habitación china" de Searle) da lugar a la introspección y reflexión crítica, que no son reductibles al funcionamiento de programas. Finalmente, en la quinta sección se argumenta que el proceso de la introspección y la reflexión crítica, son la "piedra de tope" de la IA clásica; en efecto, es la ausencia de ambas capacidades lo que impide que el niño-MT se convierta en una mente adulta pensante.

Palabras clave

Niño-máquina, mente adulta pensante, máquinas de Turing, algoritmos, introspección, inteligencia de máquina.

[*] Profesor asociado del Centro de Estudios Cognitivos y del Departamento de Filosofía, en la Facultad de Filosofía y Humanidades, de la Universidad de Chile. Sus intereses de investigación son filosofía de la mente e inteligencia artificial, ontología social y epistemología social. Ha publicado un libro titulado *Experimentos mentales y filosofías de sillón: desafíos, límites, críticas* y artículos en revistas como *Southern Journal of Philosophy*, *AI and Society*, *Anales del Seminario de Historia de la Filosofía*, *Isegoría*, *Aurora*, *UNISINOS*, entre otras. Google Académico: https://scholar.google.cl/citations?user=YBcwekAAAAAJ&hl=es
Índice h: 8

Abstract

According to the philosophers of Artificial Intelligence (AI), Turing Machines and the Imitation Game are the most important concepts proposed by Alan Turing. The Child-Machine Project, which projects learning machines via digital computers, is less known, although it is no less important. According to Turing's project, a programmed machine needs to be a Child-Machine to turn into an adult mind, one that understands, judges, and distinguishes. In this article, I argue that Turing's desideratum is not realizable only with algorithms. In the first section, I introduce the problem, while in the second I briefly analyze concepts such as algorithms, Turing Machines, and their relation. In the third section, I deal with Machine Intelligence and the Child-Machine Project. In the fourth section, I look at a form of understanding, which is the basis of the Chinese Room Argument: introspection and reflective thinking, two factors that enable the process by which results are revised. In the fifth section, I analyze why those processes of revision are the stumbling block of classical AI or GOFAI; as I argue, introspection and reflective thinking are the cognitive faculties that prevent the child-machine from becoming a "thinking adult mind".

Keywords

Child-Machine, Thinking Adult Mind, Turing Machines, Algorithms, Introspection, Machine Intelligence.

Nuestro problema es, entonces, cómo programar
una máquina para imitar al cerebro, o si lo pudiésemos
expresar de una manera más breve y menos rigurosa, para que piense.
Alan Turing, entrevista para la BBC (1951)

La educación no es aprender hechos, sino entrenar a la mente a que piense.
Albert Einstein

Introducción

La inteligencia artificial (IA) es una disciplina que tiene como fin la creación de máquinas programadas capaces de imitar la inteligencia humana. Es, entonces, una aproximación antropocéntrica a la inteligencia. Una cuestión interesante, ligada a la disciplina, es que Turing, quien es considerado uno de los padres de la IA, ideó además de la máquina que lleva su nombre y que define qué es computar, un método para establecer si es justificado atribuirles estados mentales a los computadores digitales. Dicho método está basado en un juego, el famoso y controvertido "juego de la imitación", descrito en detalle más abajo. Pero ese no fue el único aporte de Turing a la IA. Otro que es menos discutido es el proyecto del niño-máquina, es decir, la propuesta de que un programa computacional fuera capaz de aprendizaje al igual que un niño. Tal proyecto representa un desafío notable para la IA, porque consiste en proyectar un programa que aprende al igual que un

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

niño para, finalmente, convertirse en una *mente adulta pensante*. ¿En qué sentido "pensante"? En el de entender y juzgar lo verdadero y lo falso, lo claro y lo ambiguo, y lo solucionable de lo no solucionable.

En vista del optimismo de Turing, el objetivo de este artículo es justamente examinar el siguiente problema: ¿puede llevarse a cabo el proyecto del niño-máquina? Pero la pregunta que cabe es ¿por qué no podría realizarse? Como el mismo Turing señala, el niño-máquina debe convertirse en una mente adulta mediante aprendizaje. Y esta, como se sabe, incluye una serie de facultades clave para el desarrollo de la inteligencia: memoria, raciocinio, introspección, reflexión crítica, etc. Son justamente estas últimas dos capacidades las que representan un escollo crucial para el proyecto de Turing, tal como se examina aquí. En efecto, la idea que se defiende en este trabajo es que la capacidad de la mente de introspección y de reflexión crítica *no es reductible* al funcionamiento de programas computacionales, los cuales solo se basan en algoritmos con procesamiento automático de información. Justamente, el argumento que se propone es que, dado que los algoritmos son mecánicos y automáticos, y que no requieren de *insight* o de introspección alguna, no permiten la reflexión crítica, que es clave en una mente adulta pensante. Este problema es un tema actual y controversial en la IA, debido al *deep learning*. Por este motivo, aquí se emplea un método de análisis conceptual pata testear si este tipo de IA con su aprendizaje de máquina logra *insights* y reflexión crítica, que son clave para que una mente adulta aprenda efectivamente.

El artículo está dividido en cinco secciones. La primera realiza un breve recuento histórico de los algoritmos y de las máquinas de Turing; se intenta que los lectores comprendan a cabalidad qué implicancias se siguen del funcionamiento algorítmico de un programa. La segunda gira alrededor de dos problemas: por una parte, cómo Turing elaboró el concepto de *machine intelligence*, que es fundamental para entender la IA clásica y el *deep learning*; por otra, cómo su propuesta de método para testear estados mentales en máquinas programadas devino en un proyecto de aprendizaje basado en algoritmos: el proyecto del niño-máquina. La tercera parte trata con un desafío a la IA clásica —o fuerte, en términos de Searle (1980, 1990)— basado en una forma de entendimiento: la lingüística; como se argumenta con base en el contraejemplo de "la habitación china", la introspección no es reductible al funcionamiento de programas, al contrario, estos no son capaces de generarla. En cuarto lugar, se desarrolla más la idea de por qué esta capacidad y la reflexión crítica a la que da lugar, son la *piedra de tope* de la IA clásica; en efecto, sin introspección y reflexión crítica, la mente no es capaz de revisar el sentido de las reglas que se siguen automá-

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

ticamente, para así aprender realmente. Finalmente, la última sección examina las conclusiones más importantes del análisis efectuado en el trabajo.

De algoritmos a máquinas de Turing

La imitación del comportamiento inteligente es el objetivo central de la IA clásica. Con clásica refiero a GOFAI (*good old fashioned artificial intelligence*),[1] la IA que imita la inteligencia humana teniendo presente el paradigma de reglas y representaciones. Tal como lo pondría Marvin Minsky: el objetivo de la IA es la creación de máquinas programadas para realizar tareas que requieren la misma inteligencia que si fuesen hechas por seres humanos. Dichas tareas incluyen actividades simples como jugar damas, o más complejas, como la detección de COVID-19. De alguna forma, GOFAI asume que todos los problemas pueden tener un abordaje algorítmico, es decir, que pueden resolverse mediante un conjunto de pasos finitos, siendo uno de ellos recursivo. Por tanto, la IA clásica construye máquinas que sean capaces de imitar el comportamiento inteligente humano, el lingüístico, mediante procesamiento algorítmico. Pero ¿qué son los algoritmos, un concepto en boga en nuestra época?

Los algoritmos, pese a su prevalencia en el mundo de hoy, no son nuevos. Fueron popularizados por el matemático persa Abu Ja'Far Mohammed ibn Mûsâ al-Khowâzarim cerca del año 825 d. C. (Penrose, 1989, pp. 41-44), pero eran conocidos desde mucho antes. Por ejemplo, el algoritmo de Euclides, el cual consiste en un conjunto de reglas para encontrar el máximo común denominador (MCD) de dos números enteros. Dado este problema, el algoritmo tiene reglas y pasos finitos, siendo el tercer paso recursivo:

i. Dividir número y divisor, anotando resultado y remanente (R);
ii. Si R = 0, *halt*;
iii. Si R ≠ 0, tomar divisor y remanente anteriores para ejecutar paso 1.

Podemos aplicar un ejemplo de este proceso con los números 99 y 15:

Número	Número divisor	Resultado	Remanente
99	15	6	9
15	9	1	6
9	6	1	3
6	3	2	0

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

Luego, el algoritmo arroja el siguiente resultado: el MCD entre 99 y 15 es el número entero 3. Así de simple un humano puede *operar automáticamente* con el algoritmo de Euclides (vuelvo sobre la importancia del algoritmo de Euclides en la última sección).

Otro concepto que está ligado al de algoritmo es el de "descomposición recursiva", esto es, la simplificación en pasos mecánicos de una operación compleja. Por ejemplo, se puede descomponer la multiplicación recursivamente así, en la suma y la resta, respectivamente (Block, 1990, p. 256):

> M x N = A
> Las reglas de este nuevo algoritmo son las siguientes:
> i. Sumar 1 vez M a A y restar 1 a N;
> ii. Si N = 0, *halt*;
> iii. Si N ≠ 0, ejecutar paso i.

Por ejemplo, 3 x 3 puede descomponerse recursivamente en sumas y restas, hasta llegar a un resultado, algorítmicamente, que detenga el procesamiento de información.

M	x	N	=	A
3		3		0
3		2		3
3		1		6
3		0		9

Nótese que este nuevo algoritmo es una *máquina de multiplicar*: implementa un programa automáticamente y así opera recursivamente, mediante la descomposición de un problema más complejo, la multiplicación, en pasos más simples y mecánicos, la suma y la resta.

No obstante, un algoritmo no tiene exclusivamente que ver con matemáticas. Para encontrar la llave de la cerradura en un llavero, puede aplicarse un algoritmo en el que recursivamente se pase a la siguiente llave a la izquierda si no encaja la que se tiene en la mano. Y así puede operarse hasta que encuentre la llave, momento en que se detiene el proceso. Si una noche, un vecino ebrio tratara de encontrar la llave, podría ejecutar el mismo algoritmo y lo haría pese a estar semiinconsciente, porque para operar algorítmicamente no se requiere de conciencia, ni de ningún *insight*. Al contrario, estos están del todo ausentes en el procesamiento serial.

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

Justamente, los algoritmos son máquinas porque se opera con ellos de manera automática, en función de un mecanismo que es recursivo. Cabe destacar que no se requiere de esfuerzo introspectivo, entendimiento (*insight*) o ingenio para implementar un algoritmo. Por ejemplo, el vecino ebrio podría buscar la llave estando semiinconsciente (*i. e*. sonámbulo) y aun así implementar el algoritmo "encuentra la llave" de manera mecánica.

La presencia de los algoritmos es impresionante en la actualidad (*cf*. Kowalkiewicz, 2024). Están en todas partes porque son la esencia de qué es computar y de las enormes facilidades que otorga. Dicha acción procesa información, transformando *inputs* en *outputs* de manera automática y mecánica, sobre la base de un programa que incluye reglas para la transformación expresadas en fórmulas condicionales de "Si… entonces…". Luego, correr un programa consiste en la implementación de un algoritmo con capacidad de procesar información. Por esto los conceptos de algoritmo y programa computacional se traslapan. Un programa opera algorítmicamente, mientras que un algoritmo es un programa para llegar a la resolución de un problema.

Turing (1936) fue —desde un punto de vista histórico— quien precisó la definición de qué es computar. Lo hizo introduciendo una definición bajo la forma de una máquina abstracta, la denominada máquina de Turing (MT). En particular, las MT fueron postuladas como una manera de abordar el *entscheidungsproblem*, planteado por el matemático David Hillbert. En vista de dicho problema, se intenta determinar si un algoritmo X nos posibilita inferir mediante una función computable todos los teoremas de la lógica de primer orden (lenguaje formal con cuantificadores que alcanzan a variables de individuos, con predicados y funciones cuyos argumentos son constantes o variables de individuos). Esta sección no ahonda en el *entscheidungsproblem* mismo, basta decir que, gracias a este, Turing ideó las MT como dispositivos abstractos que definen de manera precisa qué es computar. Además, a raíz de dicho problema, Turing tuvo como objetivo encontrar un método para caracterizar todas las funciones computables.

Computar es, entonces, transformar un *input* en *output* sobre la base de un conjunto de reglas o un programa (y, por tanto, un algoritmo). De hecho, una MT no hace más que implementar una función computable, la cual está ligada a dos nociones fundamentales: "máquina" y "procedimiento mecánico" (se vuelve sobre estas nociones más abajo, a la luz del análisis del concepto *machine intelligence*). Una MT es justamente un dispositivo mecánico que implementa procedimientos de cálculo definibles mediante pasos finitos, es decir, implementa un algoritmo. La MT

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

posee un conjunto discreto de estados posibles y que son de número finito (aunque potencialmente enorme). Lo anterior le da a dicha máquina un número elevadísimo de cálculos posibles (aunque finito).

Los *inputs* (*i. e*. números) no tienen límite en cuanto a su tamaño y la capacidad de almacenamiento externo de la MT, donde escribe, o la cinta, que es ilimitada, al igual que los *outputs*. La MT, por definición, no internaliza los datos o cálculos externos, sino que opera con los datos o cálculos dados en las operaciones más inmediatas. Esta idea es crucial para entender por qué una MT no requiere *insights* e introspección (se trata esta cuestión nuevamente más abajo).

Tal como Penrose (1989) destaca, el tamaño ilimitado de *inputs* y *outputs*, y la capacidad de almacenamiento ilimitado de la cinta, que es infinita, dan cuenta del carácter altamente idealizado, abstracto y matemático de una MT: "Es la naturaleza ilimitada del *input*, del espacio de cálculo, y del *output* lo que indica que estamos frente a una idealización matemática en vez de algo que podría construirse en la práctica" (p. 35) (traducción propia).

De esta forma, una MT se caracteriza usualmente como una cinta infinita, con una cabeza que lee y escribe símbolos en función de un programa (las reglas del algoritmo). La cabeza "recuerda" algunos de los símbolos: está en un estado interno q_1, q_2, q_n, etc., en un momento t_n. Luego, al leer un símbolo (el *input*) y estar en un estado interno, generará un *output* en función del programa, lo cual llevará a la MT a borrar o mantener el símbolo leído, moverse y pasar a un nuevo estado, si así lo estipulan las reglas.

Por ejemplo, es posible hacer que una MT sume 3 + 2, con números enteros (Kim, 2006, pp. 125-128), dado el siguiente estado de la cinta (en esta notación el número *n* está representado por la secuencia de *n* golpes, donde cada uno ocupa un cuadrado y solo uno):

#	#	1	1	1	+	1	1	#	#
		$\uparrow q_0$							

Es importante tener en cuenta que todas las instrucciones de una MT, que es de estados discretos[2] o clics, se encuentran expresadas en una tabla de máquina. En este caso, las acciones están descritas por las reglas del siguiente programa:

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

	q_0	q_1
1	1D q_0	# Halt
+	1D q_0	
#	# I q_1	

Lo que hará la máquina es que, dado el estado interno q_0 de la cabeza, leerá *input* 1 y generará *output*: no escribir, moverse a la derecha y seguir en q_0:

#	#	1	1	1	+	1	1	#	#
			↑q_0						

Luego, leerá el siguiente 1, no escribirá, se moverá a la derecha y seguirá en q_0:

#	#	1	1	1	+	1	1	#	#
				↑q_0					

Luego, al leer el tercer 1, no escribirá, se moverá a la derecha (y seguirá en q_0), tal como vemos a continuación:

#	#	1	1	1	+	1	1	#	#
					↑q_0				

Posteriormente, sucederá algo "novedoso": leerá el *input* +, lo cambiará por un 1, se moverá a la derecha y seguirá en q_0:

#	#	1	1	1	1	1	1	#	#
						↑q_0			

Al encontrar el siguiente 1, lo dejará intacto, se moverá a la derecha y seguirá en q_0. Y hará exactamente lo mismo con el siguiente 1:

#	#	1	1	1	1	1	1	#	#
							↑q_0		

Nuevamente ocurrirá algo novedoso cuando lea # y esté en q_0. Entonces, el *output* que generará será: no escribir, moverse a la izquierda y pasar a q_1:

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

#	#	1	1	1	1	1	1	#	#
								$\uparrow q_0$	

Finalmente, estando en q_1 y leyendo 1, escribirá # y hará *halt,* tal como se ilustra:

#	#	1	1	1	1	1	#	#	#
							$\uparrow q_0$		

En consecuencia, la MT procesa algorítmicamente, en función de las reglas de un programa, indicadas en la tabla de máquina, con las reglas de esta, que se expresan en condicionales de la forma "si… entonces…". Y así, supuestamente, resuelve todos los problemas que requieren solución algorítmica. Es importante notar que todos los problemas, según Turing, pueden tener un abordaje algorítmico, incluso cuando se requiere de aprendizaje. Esto justamente lo inspira a concebir el proyecto del niño-máquina, tal como se examina en la siguiente sección. Este proyecto justamente nace como consecuencia de eliminar la pregunta "¿pueden pensar las máquinas?" (Turing, 1950).

Machine intelligence y el proyecto del niño-máquina de Turing

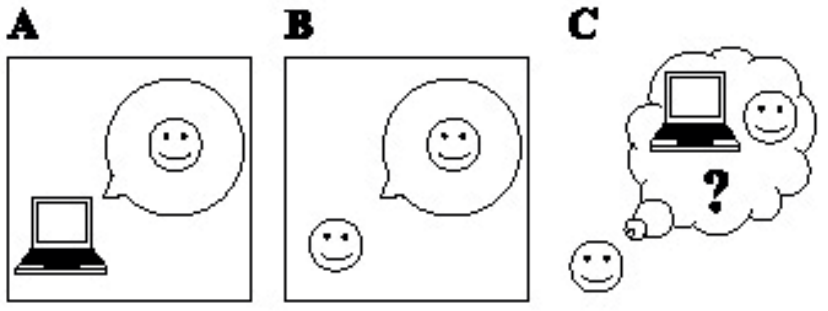
Pese a su presencia en la actualidad, el concepto *machine intelligence* o inteligencia de máquina, fue originalmente concebido por Alan Turing en los años 50. Teniendo presente el problema del entendimiento lingüístico y de cómo representaba una llave para adjudicar estados mentales a máquinas, Turing ideó una manera de evadir la pregunta "¿pueden pensar las máquinas?". Lo hizo por la siguiente razón: los términos "pensar" y "máquina" pueden generar disenso, ante los usos alternativos de las personas, lo cual puede llevar a una suerte de encuesta tipo Gallup. En efecto, Turing quiere alejarse de las definiciones, por llevar a los usos de los conceptos.

Como una manera de evadir el análisis de los términos "pensar" y "máquina", Turing propuso un método empírico que permitiera |recabar evidencia de la existencia de estados mentales en máquinas programadas: el juego de la imitación. Pese a que existen al menos dos versiones del juego,[3] la tradición ha interpretado que puede describirse este en una especie de versión estándar, tal como la figura 1 ilustra. En esta versión, el juego consiste en que una máquina programada se hace pasar por una persona,

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

engañando interrogadores (Saygin *et al*., 2000), mientras que una persona responde sinceramente desde una segunda habitación. Los jueces o interrogadores, luego de rondas de preguntas de cinco minutos, tienen como misión discernir si están en presencia de una máquina o de una persona, solo con las respuestas tipeadas.

Figura 1
Ilustración estándar del "juego de la imitación"



Ahora Turing describe de manera precisa: quiénes pueden ser interrogadores y qué tipo de preguntas pueden formularse. En relación con el primer punto, sostiene que los interrogadores tienen que ser "promedio", es decir, no pueden ser especialistas en ciencias de la computación o similares, porque ello les daría una ventaja respecto de descubrir a la máquina. Con relación al segundo punto, precisa que las preguntas tienen que ser promedio también, de modo de no darle una ventaja evidente a los seres humanos. En consecuencia, los interrogadores y las máquinas deben ser promedio de modo de no dar ventajas a los humanos en el descubrimiento de las máquinas. Pero ¿cuáles pueden participar en el juego?

Turing establece que las máquinas en cuestión son los computadores digitales (máquinas programadas). Estas tienen la arquitectura básica de procesamiento, unidad y control en su funcionamiento. Además, son digitales porque operan mediante la manipulación de símbolos de acuerdo con las reglas de un programa. Es decir, los computadores digitales procesan la información sintácticamente, dado los algoritmos que implementan (se vuelve sobre la importancia de la sintaxis en la manipulación simbólica más abajo).

Teniendo presente el juego de la imitación —en su versión estándar— y los participantes en el mismo, Turing (1950) hizo una predicción:

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

> Creo que en alrededor de cincuenta años será posible programar un computador, con una capacidad de memoria de 10^9, para que participe en el Juego de la Imitación tan eficientemente que un interrogador lego no tendrá más de un 70 % de probabilidad de hacer la identificación correcta después de cinco minutos de interrogatorio. Creo que la pregunta original, ¿pueden las máquinas pensar?, es demasiado absurda para seguir analizándola (p. 49) (traducción propia).

Por supuesto, la predicción de Turing es tan polémica como su test y lo es por dos razones importantes. En primer lugar, no se hace cargo de que el contexto de la predicción involucra a la filosofía de la mente y de la IA. En segundo lugar, la disciplina de la filosofía de la IA se inicia con una particular descripción de Babbage (en Swade, 2000) de una de sus máquinas:

> Babbage habla de *enseñarle a la máquina a prever* […] En otras ocasiones afirma que la máquina *sabe*. […] La analogía entre estos actos y los procesos mentales me forzó al uso figurativo de tales términos. El uso de estos fue ponderado como económico y expresivo, y prefiero seguir usándolos que sustituirlos por largos circunloquios (pp. 103-104) (traducción propia) (énfasis añadido).

Babbage duda que a su máquina puedan adjudicarse estados mentales de manera totalmente literal. En vez de eso, propone que dicha adjudicación se hace por economía conceptual, o como una forma de evitar largos circunloquios. De esta manera, la predicción de Turing queda relativizada en el ámbito de la reflexión filosófica, a saber, aquella que intenta examinar y argumentar con relación a la existencia de estados mentales en máquinas programadas. Pero Turing (1950) enfrentó dos objeciones aún más serias, tal como se detalla a continuación:

> ¿Podría la máquina hacer algo diferente de lo que hace el hombre? Esta objeción es muy poderosa, pero podemos decir que, no obstante, si una máquina puede ser construida para jugar el juego de la imitación satisfactoriamente, no necesitamos hacernos cargo de ella (p. 42) (traducción propia).

Procede entonces a preguntarse lo siguiente, con una objeción que tiene ribetes filosóficos importantes por el tipo de situación planteada, claramente hipotética: "No estamos preguntando si los computadores digitales pueden tener un buen desempeño en el juego, o si hay en el tiempo presente tales máquinas, sino *si hay computadores imaginables que pudieran tener buen desempeño*" (p. 43) (traducción propia) (énfasis original).

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

Como puede apreciarse, Turing propone una suerte de experimento mental en relación con la capacidad intelectual de los computadores digitales. Debe haber *computadores digitales posibles* que puedan pasar el test. Si, de una manera optimista, se asume que sí puede haber tales computadores digitales, la pregunta que cabe es: ¿en qué sentido deben ser inteligentes respondiendo a las preguntas?

Luego de anticipar una serie de objeciones posibles a la predicción,[4] Turing concluye que la mejor forma de llegar a concebir un computador digital capaz de tener un buen desempeño en el juego de la imitación es mediante el aprendizaje. Pero no cualquier aprendizaje, sino el de un niño-máquina. Con esto, Turing pone en el tapete que es mediante la educación y el aprendizaje que se logrará llegar a tener máquinas programadas que puedan *imitar el comportamiento inteligente humano*. ¿En qué sentido de "imitar"?

Una cuestión que es importante esclarecer es el tipo de funcionalismo al que finalmente adscribe Turing (Putnam, 1967, 1968). En vista del juego de la imitación, no es necesario replicar las propiedades físicas o biológicas del cerebro, pues el juego permite diferenciar las capacidades mentales de dichas propiedades.[5] De acuerdo con este tipo de funcionalismo de MT, es mejor imitar la mente de un niño, ya que es como un libro de notas, con mecanismos simples (programables) y páginas en blanco. Esto es clave para entender de qué forma Turing sostiene que la mente puede mecanizarse y que cualquier imitación exitosa de la máquina-cerebro replicará este y su capacidad inteligente. De esta forma, la mente de un niño-máquina se puede transformar en una mente adulta y lo hará en términos de cómo aprende quien piensa, razona y entiende (Feldman, 2009, pp. 70-72).

Las palabras de Turing (1950) en relación con su proyecto del niño-máquina son las siguientes:

> En el proceso de tratar de imitar a una mente humana estamos condicionados a pensar bastante acerca de los procesos de que fue sujeta para llegar a su estado actual. Podemos notar tres componentes:
>
> 1. El estado inicial de la mente en el nacimiento;
>
> 2. La educación que ha recibido;
>
> 3. Otra experiencia, no descriptible como educación, que ha tenido.

Turing (1950) propone, entonces, de la siguiente forma un tipo específico de aprendizaje, para imitar qué hace una mente adulta. Para esos

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

efectos, incluso asume que la mente de un niño tiene mecanismos que pueden ser, a todas luces, algorítmicos:

> En vez de intentar producir un programa para simular la mente humana, ¿por qué no tratamos más bien de simular uno que sea como la de un niño? Si se le sometiera a la educación apropiada, uno obtendría una mente adulta. Tal vez el cerebro del niño es algo así como un libro de notas que uno compra en una tienda de artículos de oficina. Pequeños mecanismos y un montón de páginas en blanco (p. 62) (traducción propia).

La pregunta que cabe, no obstante, es si el proyecto del niño-máquina es viable como Turing lo concibió originalmente. Justamente, en las siguientes secciones se muestra que, dado que hay partes del aprendizaje que dependen de la introspección, el proyecto del niño-máquina no es realizable al modo pensado por Turing (*i. e.* solo mediante algoritmos). Es decir, hay dificultades en principio para que un niño-máquina se transforme en una mente adulta, que es justamente lo que Turing propuso para concretar el proyecto de IA de largo plazo.

"La habitación china" y su puzle cartesiano

Una cuestión que vale la pena enfatizar es que, además del conductismo que se le achaca, al test de Turing subyace una suerte de cartesianismo. Ello ocurre porque, como se sabe, Descartes (1994, pp. 112-113) propuso que hay dos signos inequívocos de inteligencia: el uso de signos convencionales lingüísticos y la acción inteligente, guiada por razones. En relación con el primer punto, el filósofo francés establece que solo el ser humano es capaz de usar los signos convencionales, ya que los animales solo son capaces de usar signos naturales. Por ejemplo, Juan es capaz de enunciar "amo a María" o bien "María es amada por mí", dos oraciones que significan lo mismo. En cambio, los animales solo son capaces de reaccionar a los estímulos de manera natural. Si a un perro se le pincha una pata, aullará de dolor y lo hará conminado por su propia naturaleza, es decir, por la disposición de sus órganos. Por otra parte, Juan puede dar razones para que se entienda por qué ama a María, en cambio, Bobby, el perro de Juan, solo actúa en función de la propensión de sus órganos cuando aúlla de dolor.

Turing hereda un prejuicio cartesiano, a saber, que el uso y manejo del lenguaje es signo de inteligencia (Descartes, 1994, pp. 112-113). En este sentido, el juego de la imitación no es más que una secuela de las ideas de Descartes en relación con la razón y la inteligencia. En efecto,

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

el hecho de que el computador digital, con la adecuada programación, almacenamiento y rapidez, sea capaz de responder como lo haría una persona, engañando interrogadores, indica que el lenguaje o la conducta lingüística es aquello que permite *verificar la existencia de inteligencia*. Nótese que, pese a que Turing sostiene que *solo debe haber computadores posibles* que pasen el test, hacerlo indica que estamos en presencia de un test ácido de la inteligencia, al menos respecto de la existencia de estados mentales en máquinas programadas.

Pero hay otro elemento cartesiano que debe remarcarse y es la cuestión del entendimiento lingüístico. Este solo puede detectarse de manera interna, por ponerlo de algún modo. John Searle (1980) es enfático cuando habla de su experimento mental contra la GOFAI, que él llama "IA fuerte" (p. 67). El experimento mental de la habitación china está pensado contra la tesis de que la mente es un programa computacional o *software*. En la habitación, Searle es un hablante nativo de inglés, que habla cierto grado de francés y nada de chino. ¿Cómo lo sabe? Internamente: si se presenta un inglés, le entenderá todo lo que dice; si se presenta un francés, entenderá parte de lo que dice; si se presenta un chino hablándole, no le entenderá absolutamente nada. En consecuencia, se examina y determina si se entiende un lenguaje de manera interna, gracias a la introspección (se vuelve sobre este punto más abajo). ¿Cómo podrían las máquinas programadas entender un idioma, entonces?

Dos investigadores de la GOFAI, Roger Schank y Robert Abelson (1977), concibieron el entendimiento lingüístico de una manera anticartesiana, esto es, apelando a la descomposición recursiva de elementos. Tal como se recordará, esta va de lo complejo a lo más simple, hasta alcanzar un nivel de simplicidad mecánica. Por ejemplo, ¿qué se hace cuando hay que entender una palabra? Entender una palabra es posible debido a que hay tres etapas claramente diferenciadas: la de traerla, la de cotejarla con un listado para hacerla calzar con alguna otra, y la de recuperar la información sintáctica y semántica asociada. Es importante señalar que estos tres niveles son en sí mecánicos, porque traen, cotejan y recuperan información, respectivamente. De acuerdo con este modelo de entender lingüísticamente, no hay una suerte de *yo* cartesiano que piense y corrobore lo que la palabra significa. Por el contrario, hay solo pasos mecánicos, algorítmicos, que procesan información.

Ahora bien, la teoría de Schank y Abelson puede ser extrapolada al entendimiento de historias y de responder preguntas acerca de información que no está explícitamente desarrollada en ellas. Tal cuestión la hace un computador programado: SAM (*script applier mechanism*).[6] El sistema

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

opera de una manera similar a entender una palabra: hay una cantidad de *scripts* o historias almacenadas en la memoria, hay historias que entran, se hacen calzar estas con las primeras, y luego se puede responder preguntas de información que no está explicitada. Una cuestión digna de destacar es que una versión ideal de SAM ciertamente pasaría el test de Turing. Es decir, en función de los *scripts*, y de las historias, el computador podría ser capaz de proporcionar información que no está directamente aludida en ellas, lo cual ciertamente cuenta como comportamiento inteligente. Sin embargo, Searle no piensa que SAM entiende historias y es inteligente. Eso solo ocurriría si la IA fuerte fuera verdadera.

En función de lo defendido por Schank y Abelson (1977), Searle (1980) se empeña en mostrar que la IA fuerte es una teoría falsa. Una manera de mostrar la falsedad de una teoría es preguntando qué sucedería si la mente operase de acuerdo con ella, por ejemplo, qué acontecería si la mente operase de acuerdo con los postulados de la IA fuerte; en consecuencia, propone falsarla indicando qué sucedería si la mente operase de acuerdo con esta aproximación teórica.

Así volvemos al escenario del experimento mental. Como se sabe, en chino no hay abecedario, sino ideogramas, es decir, representaciones pictóricas de eventos, cosas, etc. Los hablantes de chino saben del significado de los ideogramas en virtud de su forma, cuestión esencial al experimento mental y al hecho de que computar es manipular símbolos con base en dicha forma y reglas. El hablante está encerrado en una habitación, la cual posee una rendija de *inputs*, una rendija de *outputs*, un banco de datos con *scripts* en chino (que él solo ve como símbolos sin sentido) y un libro de reglas para manipular los símbolos. Ahora, hay una serie de hablantes nativos de chino fuera de la habitación que mandan ideogramas por la rendija de *input*. El sujeto toma esos ideogramas, los compara con el banco de datos y procede a manipular dichos símbolos en virtud de su forma, gracias al libro de reglas, que está escrito en inglés. Dicho libro estipula que si, por ejemplo, los ideogramas 81, 99 y 100 están juntos, el sujeto debe mandar por la rendija de *output* los ideogramas 1 y 7. Y así sucesivamente con todos los *inputs*, para convertirlos en *outputs*.

Por supuesto, el sujeto no tiene idea de qué está haciendo con los ideogramas, salvo manipularlos. De hecho, solo manipula símbolos sintácticamente, en virtud de su forma, y gracias a las reglas del libro. No comprende, entonces, qué significan dichos ideogramas, y menos aún entiende que los hablantes nativos de chino le envían historias y preguntas, para obtener respuestas a través de la rendija de *output*. Los hablantes incluso podrían estar insultando al sujeto del experimento y este no se

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

percataría en lo más mínimo. Una cosa es clara, dado el escenario de la habitación china, el individuo no tiene entendimiento lingüístico del chino de manera absoluta, ni hay estados mentales asociados. Solo manipula símbolos sin saber lo que estos significan y pese a que los hablantes de chino creen que hay otro hablante de chino encerrado en la habitación. Es decir, los símbolos son solo formas manipuladas en virtud de las reglas del libro, en el caso de Searle, quien no entiende a qué va tanta manipulación simbólica. El lector atento recordará que el escenario descrito es análogo al test de Turing, pero hay una diferencia importante con este. En el caso del juego de la imitación, la predicción de Turing es que en el año 2000 el interrogador promedio no tendrá más de un 70 % de chance de descubrir al computador. Sin embargo, en la habitación china todos los interrogadores (*i. e.* los hablantes nativos de chino fuera de la habitación) resultan engañados.

Hay una serie de objeciones al experimento mental de la habitación china.[7] Dos son las más populares y seductoras. La primera de ellas es que Searle es solo parte de un sistema. La totalidad de este posee entendimiento lingüístico del chino, a pesar de que no hay un *locus* claro de dicho entendimiento. Searle, entonces, no puede afirmar que la totalidad del sistema no tiene entendimiento lingüístico, ya que es probable que la habitación, más las rendijas, más el banco de datos, más el libro de reglas sean capaces de entender chino. Este filósofo se defiende de la objeción del sistema haciendo hincapié en que, si la habitación entendiese, toda clase de subsistemas podrían tener estados mentales, sin que lo supiéramos. Incluso, asevera que él podría internalizar todos los elementos de la habitación, como la rendija de *input*, la de *output*, el banco de datos y el libro de reglas (que podría memorizar). Todo el procesamiento simbólico podría, entonces, realizarse internamente. Dada la importancia que tiene este punto, vuelvo sobre él más abajo.

La segunda objeción importante es la del robot. Según dicha objeción si hubiera un robot que anclara causalmente los símbolos en el ambiente gracias al uso de transductores, aquel podría entender el significado de los símbolos chinos. Claramente, esta objeción añade un elemento novedoso: que los símbolos tienen significado en la medida que hay un anclaje de los mismos en el ambiente. Searle responde a esta objeción con un nuevo experimento mental: ahora está en la cabeza del robot, recibiendo símbolos chinos desde los transductores, y enviando símbolos chinos a los elementos móviles del robot, de modo de producir la respuesta adecuada. Sorprendentemente, se vuelve a replicar el escenario de la habitación china, porque recibe símbolos cuyo significado desconoce,

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

y envía símbolos, cuyos significados también ignora. En consecuencia, el robot no parece un argumento suficientemente contundente, al menos en lo que respecta al entendimiento lingüístico.

A propósito de este, es claro que tanto el sistema como el robot se fundamentan en un elemento cartesiano. Piénsese en el escenario de Searle internalizando los elementos del sistema. Sin embargo, hay un solo elemento que no puede internalizar: él mismo. Es decir, gracias a la introspección puede dar cuenta de que entiende inglés y no chino. En otras palabras, si Searle internalizase todos los elementos del sistema, habría uno solo que no podría internalizar, a saber, él mismo que es quien ejecuta el experimento. Es por esto que la Habitación China tiene, pese a Searle, un sesgo cartesiano (González, 2012). Lo tiene porque, como la mayoría de los experimentos mentales acerca de la naturaleza de la mente, es la introspección la encargada de indicar de qué manera lo descrito por una teoría es verdadero o falso. En el caso de la habitación china, la introspección mostraría, según Searle, de qué forma la IA fuerte es falsa.[8]

La introspección no puede reducirse al funcionamiento de un algoritmo. Este elemento cartesiano de la introspección involucra un sesgo que no puede ser caracterizado algorítmicamente, tal como se examina a continuación, y que está ligado a la reflexión crítica.

La introspección y la reflexión crítica son las "piedras de tope" de la IA clásica

En esta sección final, se muestra en qué sentido la introspección no puede ser reducida en principio al funcionamiento de algoritmos. Si esto es así, el proyecto de aprendizaje del niño-MT tambalea. Lo hace porque hay una buena porción de dicho aprendizaje que, análogamente al entendimiento lingüístico, depende de la introspección, la que lleva a darse cuenta de que un proceso algorítmico es, por ejemplo, erróneo para alcanzar un resultado. Y al revés también ocurre con los algoritmos: que para llegar a un resultado no es necesario tener un *insight*, o introspección. Recuérdese el algoritmo de Euclides: para determinar el MCD de 99 y 15 se siguieron una serie de pasos, de manera mecánica, hasta llegar al resultado, 3 es el MCD de ambos números. Si bien se podría argüir que la máquina podría comprender que 3 es el MCD de 99 y 15, tal suposición sería dudosa si se supone un elemento adicional: que la máquina podría emplear un algoritmo fallido, con reglas que llevan a un *loop*, a recursión sin detención, al intentar encontrar diversos MCD con reglas erróneas. En tal caso, no

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

se podría decir que la máquina entiende el resultado. Luego, existe una diferencia crucial entre entender un problema y solución, y no hacerlo.

Por ejemplo, si a una máquina programada se le instruyera a seguir reglas finitas para encontrar un número impar mediante la suma de dos números impares, la máquina programada no daría con la solución al problema, porque este simplemente no tiene solución. En efecto, ¿comprendería la máquina que el *loop* sin detención es producto de un algoritmo que busca la solución a un problema que no la tiene? Descubrir tamaña dificultad es parte de un proceso en que participa la introspección y, que es necesaria, para no seguir reglas sin sentido. En vista de esta dificultad, uno podría concluir que los algoritmos son demasiado seriales y lineales, y que como no recurren a un proceso de introspección, que facilita la reflexión crítica, no son capaces de establecer que hay algunos seguimientos de reglas que simplemente carecen de sentido.

El automatismo de los pasos algorítmicos no lleva a ninguna experiencia psicológica consciente, tal como acontece con la habitación china. En el caso de 99 y 15, y de su máximo común denominador, no hay experiencia psicológica en relación con las matemáticas. De hecho, el algoritmo podría ser seguido por alguien ignorante en matemáticas, como también por un experto en ellas. Ninguno de los dos tendría una experiencia psicológica interesante asociada con la ejecución del algoritmo, y por esta razón no se producen estados mentales conscientes en ningún sentido. Es decir, ningún agente que siga las reglas del algoritmo de Euclides tendrá una experiencia psicológica consciente asociada al seguimiento de dichas reglas. Lo mismo, *mutatis mutandis*, sucede con el algoritmo para obtener un número impar mediante la suma de dos números impares. Tal como no hay estados conscientes asociados a la ejecución del algoritmo de Euclides, no hay estados conscientes asociados a la ejecución de un algoritmo fallido, lo que es clave para que la máquina no entienda que el problema es un sinsentido.

Tal como Penrose (1989, pp. 141-143) destaca, el *halting problem* es insoluble si no hay introspección y experiencia consciente asociadas a *insights* matemáticos (intuiciones matemáticas) que indiquen si, por ejemplo, un problema matemático no tiene solución. En cierta medida, la IA clásica paga un alto precio, a causa de la naturaleza de los algoritmos. Como estos no requieren de *insights* o de introspección, se siguen todas las consecuencias negativas en relación con la resolución de problemas cuya solución no tiene un abordaje algorítmico. Con relación a lo examinado en esta sección, resulta que el niño-máquina no podría convertirse en una mente adulta, pues gracias a la introspección, solamente esta es capaz de

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

hacer que un agente se percate si un algoritmo se detendrá. Es decir, el niño-MT, que carece de introspección, no logrará percatarse y entender de la no solubilidad de algunos problemas matemáticos, lo cual hace que no pueda convertirse en una mente adulta consciente, esto es, capaz de reflexión crítica y de comprensión frente al sinsentido de un problema.

Tal reflexión crítica sigue un patrón muy similar a la reflexión socrática, clave para ciertos procesos educativos, por ser una búsqueda constante de la verdad mediante la ironía y el *elenchos* (*i. e.* las preguntas refutatorias que adquieren sentido solo en un contexto de ciertas afirmaciones). Del mismo modo que la reflexión crítica en el caso de los algoritmos, el *elenchos* ilumina la reflexión crítica típica de un adulto, que duda, que entiende, que afirma, y que es clave para aplicar el método socrático, que cuestiona el sentido de algunas afirmaciones. A diferencia de lo que cree Corballis (2007), con su tesis acerca de la recursión como característica única de la especie humana, el proceso educativo que lleva a una mente adulta pensante no puede depender solamente en dicha recursión. Esta conduce a la cognición automática y serial, que justamente carece de reflexión crítica, y del proceso típico de la mente adulta que duda y vuelve sobre sí misma mediante la introspección. De hecho, ese proceso cognitivo puede relacionarse con la capacidad imaginativa, interpretativa y generativa de pensamiento, que pese a Corballis, no es producto de un proceso mecánico. Pensar es, en un sentido socrático, cuestionar, imaginando e interpretando contextualmente. Justamente en relación con la introspección y la reflexión crítica típica de una mente pensante adulta que se defiende aquí, Bailin y Siegel (2002) destacan que:

> El pensamiento que está dirigido primordialmente por la evaluación o crítica de ideas o productos *no es algorítmico, sino que tiene un componente generativo e imaginativo.* La aplicación de criterios no es un proceso mecánico, sino que involucra la interpretación de las circunstancias, y un juicio imaginativo en relación con la aplicabilidad de criterios en diferentes circunstancias, y con si los criterios se satisfacen (p. 187) (traducción propia) (énfasis añadido).

Ciertamente, el llamado método socrático está íntimamente ligado con esta dimensión generativa del pensamiento, imaginativa, que justamente consiste en no seguir reglas de manera automática, y en cuestionar el sentido o sin sentido de algunas de ellas. Incluso, desde el punto de vista de la evidencia empírica, hay estudios que sugieren que la reflexión crítica típica de los adultos es clave en relación con resultados positivos en la educación superior. Es decir, pese a Turing, hay eviden-

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

cia empírica que muestra cómo los humanos se convierten en mentes adultas pensantes, y cómo esta característica única de la mente humana permite logros académicos. De hecho, hay estudios que justamente exploran el nexo que existe entre el pensamiento reflexivo, el crítico, el automonitoreo, y los resultados positivos en la educación universitaria (Ghanizadeh, 2017). Otro estudio llega a conclusiones similares: el pensamiento crítico, reflexivo y creativo es fundamental para alcanzar logros académicos (Akpur, 2020).

En síntesis, los algoritmos potencian el aprendizaje, pero también tienen limitaciones importantes. Al no requerir de introspección, hacen que la resolución de problemas asociados a esta simplemente no exista. Un algoritmo antiguo y simple, como el de Euclides, muestra de manera precisa cómo la introspección está del todo ausente en el procesamiento de información, con todas las consecuencias negativas que se siguen de ello. Las preguntas que caben son: ¿Tendría Turing el mismo entusiasmo respecto del proyecto de aprendizaje del niño-máquina si se hubiese percatado de la limitación de los algoritmos? ¿Habría defendido la idea de que el niño-máquina puede aprender con base en ellos y convertirse en una mente-adulta, sin la capacidad de reflexión crítica? Una respuesta intuitiva a ambas interrogantes es que el filósofo y matemático tal vez no hubiera defendido con tanta pasión la posibilidad de que las máquinas programadas pueden aprender sin desarrollar una especie de reflexión socrática. O, al menos, que un niño-máquina puede finalmente convertirse en una genuina mente adulta aprendiendo *solo con la ayuda de algoritmos*, que es la finalidad de su controvertido proyecto de aprendizaje en la IA.

Conclusión

En este artículo se ha desarrollado un problema insoslayable para la IA clásica o GOFAI: que el procesamiento algorítmico deja fuera la introspección y la reflexión crítica. En particular, se ha examinado de qué manera se siguen consecuencias negativas de la ausencia de ambos procesos cognitivos, que son clave para que la mente de un niño se transforme en una auténtica mente adulta pensante.

Para mostrar los problemas de la IA clásica o GOFAI, se ha descrito el funcionamiento de algoritmos simples, como el de Euclides, o de algunos que no tienen detención, tal como el mencionado de la búsqueda de un número impar mediante la suma de dos números impares. En am-

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

bos se refrenda lo afirmado más arriba: que la ausencia de introspección y reflexión crítica representa un escollo para el aprendizaje genuino, especialmente al modo del proyecto de niño-MT. En efecto, esta carencia limita lo proyectado por él, particularmente en lo que respecta a cómo una mente puede no seguir reglas de manera automática.

Para llegar a tal conclusión, se han desarrollado cinco secciones. La primera consistió en presentar el problema a examinar en el ensayo. La segunda, en cambio, hizo un recuento histórico de los algoritmos y su relación con las MT. La tercera sección abordó el concepto de *machine learning*, tal como Turing la entiende, y de qué manera dicho concepto se relaciona con el proyecto de aprendizaje del niño-máquina. En la cuarta sección, se expuso el problema de la habitación china y de qué forma dicha habitación deja fuera el entendimiento, la introspección y la reflexión crítica, todos procesos cognitivos clave para el aprendizaje. En la quinta y última sección, se abordó por qué dichos procesos son la piedra de tope de la IA clásica. En particular, se examinó en qué sentido los algoritmos dejan fuera la introspección, lo cual trae consecuencias negativas para el proyecto de aprendizaje del niño-MT. En efecto, dicho proyecto queda limitado al puro procesamiento de información y, en consecuencia, a las dificultades para que el niño-máquina se transforme en una mente adulta pensante.

Siguiendo directrices similares a Weizenbaum (1984) y Smith (2019), quienes exploran las falencias de la IA en términos de desarrollar la habilidad de juzgar sin compromiso ético y sin acción responsable, aquí se ha criticado cómo el proyecto del niño-MT responde a *una concepción de la inteligencia exclusivamente algorítmica*, es decir, que solo se apoya en el funcionamiento de una máquina programable que calcula y que procesa información mediante la manipulación de símbolos mediante las reglas de un programa. Por su carácter algorítmico, la introspección queda fuera, y dificultades como el *halting problem* quedan sin solución. También quedan en entredicho todos los problemas que se solucionan apelando a la introspección y al pensamiento reflexivo, imaginativo y crítico. Cabe preguntar, entonces, ¿puede entonces la mente adulta pensante ser mecanizable en términos puramente algorítmicos, como Turing pretende? ¿Puede llegarse a tal mente pensante mediante un niño que solo aprende con base en algoritmos? Aquí la respuesta ha sido negativa: el proyecto de niño-máquina no es factible tal como Turing lo concibió, i.e., solo en función de algoritmos, pues estos son incapaces de generar introspección y pensamiento reflexivo-crítico-imaginativo. El proyecto de aprendizaje algorítmico carece de estos procesos fundamentales en la

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

educación, pese a la decidida defensa del filósofo-matemático Alan Turing de su niño-máquina.

Notas

1 En este trabajo se usa de manera intercambiable GOFAI e IA clásica. Más adelante, se refiere a como el cognitivismo se plasma en lo que Searle denomina IA fuerte. De alguna forma, todos estos términos significan lo mismo, porque la asunción básica es que la mente es, como cuestión de hecho, un computador programado de procesamiento serial y algorítmico. Es decir, GOFAI, IA clásica, el cognitivismo y la IA fuerte se apoyan en una teoría, a saber, la teoría computacional de la mente (tal como la describe Block, 1990).

2 Con estados discretos se quiere decir que una MT no puede estar en grados, sino en estados limitados, precisos y definibles. Por ejemplo, 1,5 es gradual entre 1 y 2, mientras que 0 y 1 son estados discretos.

3 Para un examen más acucioso (*cf*. González, 2015). En este ensayo se muestra por qué la identificación del sexo de los participantes no es casual en el Juego de la Imitación: para tener el comportamiento lingüístico femenino no se requiere tener el cerebro femenino, lo cual muestra la crucial diferencia entre las propiedades físicas de este y sus capacidades intelectuales.

4 Turing adelanta nueve objeciones posibles a su Juego de la Imitación: la teológica, la de las cabezas en la arena, la matemática, el argumento de la conciencia, el argumento de las múltiples discapacidades, la de Lady Lovelace, la de la continuidad del sistema nervioso, la de la informalidad de la conducta y la de percepción extrasensorial (Turing, 1950, pp. 49-60). Por razones de espacio aquí solo se mencionan dichas objeciones.

5 Turing, pese a lo comentado por algunos (*i. e*. Block, 1990, pp. 248-253) no es un conductista, sino un funcionalista. Se pueden revisar cómo el funcionalismo MT es antibiológico (Putnam, 1967, 1968; Block, 1990, 1995; Heil, 2004; González, 2011).

6 Para una revisión de cómo trabaja SAM puede revisarse el funcionamiento de otro *chatbot*: ELIZA (Weizenbaum, 1984).

7 *Cf*. Block (1995) y Preston y Bishop (2002).

8 Por razones de espacio solo se consigna que, dadas las objeciones al experimento mental, la habitación china plantea una duda razonable de que la IA fuerte es verdadera. Pero no resulta tan claro que el experimento mental de Searle refute de manera definitiva la IA fuerte. De esta forma, al rebajar el resultado del experimento mental a duda razonable en vez de refutación se salvan las objeciones del sistema, del robot, entre muchas otras.

Bibliografía

AKPUR, Uğur
2020 Critical, reflective, creative thinking and their reflections on academic achievement. *Thinking Skills and Creativity*, *37*. https://doi.org/10.1016/j.tsc.2020.100683

Sophia 39: 2025.

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

BAILIN, Sharon, & HARVEY, Siegel
2003 Critical thinking. En N. Blake, P. Smeyers, R. Smith, & P. Standish (eds.), *The Blackwell Guide to the Philosophy of Education* (pp. 181-192). Oxford: Blackwell.
BLOCK, Ned
1990 The computer model of the mind. En D. N. Osherson y E. E. Smith (eds.), *Thinking: An Invitation to Cognitive Science* (vol. 3, pp. 247-289). Cambridge, MA: MIT Press.
BLOCK, Ned
1995 The mind as software of the brain. En J. Heil (ed.), *Philosophy of Mind: A Guide and Anthology* (pp. 267-274) Oxford: OUP.
CORBALLIS, Michael
2007 The Uniqueness of human recursive thinking: The ability to think about thinking may be the critical attribute that distinguishes us from all other species. *American Scientist*, *95* (3), 240-248. https://doi.org/10.1511/2007.65.240
DESCARTES, René
1994 *Discurso del método*. Madrid: Alianza.
FELDMAN, Richard
2009 Thinking, reasoning and education. En H. Siegel (ed.), *The Oxford Handbook of Philosophy and Education* (pp. 67-82). Oxford: OUP.
GHANIZADEH, Afsaneh
2017 The interplay between reflective thinking, critical thinking, self-monitoring, and academic achievement in higher education. *Higher Education*, *74*(1), 101-114. https://doi.org/10.1007/s10734-016-0031-y
GONZÁLEZ, Rodrigo
2011 Máquinas sin engranajes, cuerpos sin mente. *Revista de Filosofía Universidad de Chile*, *67*, 183-200. http://dx.doi.org/10.4067/S0718-43602011000100012
2012 La pieza china: un experimento mental con sesgo cartesiano. *Revista Chilena de Neuropsicología*, *7*(1), 1-6. https://bit.ly/4kZC7Wb
2015 ¿Importa la determinación del sexo en el test de Turing? *Aurora*, *27*(40), 277-295. http://dx.doi.org/10.7213/aurora.27.040.AO02
KIM, Jaegwon
2006 *Philosophy of Mind*. Cambridge, MA: Perseus Books.
KOWALKIEWICZ, Marek
2024 *The Economy of Algorithms: AI and the Rise of the Digital Minions*. Bristol: Bristol University Press. https://doi.org/10.2307/jj.10354686.14
HEIL, John
2004 Functionalism. En *Philosophy of Mind: A Guide and Anthology* (pp. 139-49). Oxford: OUP.
PENROSE, Roger
1989 *The Emperor's New Mind*. Oxford: OUP.
PRESTON, John, & BISHOP, Michael
2002 *Views into the Chinese Room*. Oxford: OUP.
PUTNAM, Hilary
1967 Psychological predicates. En J. Heil (ed.), *Philosophy of Mind: A Guide and Anthology* (pp. 158-167). Oxford: OUP.
1968 Brains and behaviour. En D. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 45-54). Oxford: OUP.

Sophia 39: 2025.

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

SAYGIN, Ayse Pinar, CICEKLI, Ilyas, & AKMAN, Varol
2000 Turing test: 50 years later. En J. H. Moor (ed.), *The Turing test: The Elusive Standard of Artificial Intelligence* (pp. 23-78). Dordrecht: Kluwer Academic Publishers.
SCHANK, Roger, & ABELSON, Robert
1977 *Scripts, Plans, Goals, and Understanding*. Hillsdale, NJ: Erlbaum.
SEARLE, John
1980 Minds, brains and programs. *The Behavioral and Brain Sciences*, *3*, 417-457. https://bit.ly/45XNcTg
1990 Is the brain's mind a computer program? *Scientific American*, 1 de enero, 26-31. https://bit.ly/3FS93kA
SMITH, Brian
2019 *The Promise of Artificial Intelligence: Reckoning and Judgement.* Cambridge, MA: MIT Press.
SWADE, Doron
2000 *The Difference Engine: Charles Babbage and the Quest to build the First Computer*. Londres: Penguin.
TURING, Alan
1936 On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230-265. https://doi.org/10.1112/plms/s2-42.1.230
1950 Computing intelligence and machinery. *Mind*, *59*(236), 433-60. https://bit.ly/3ZXWafy
WEIZENBAUM, Joseph
1984 *Computer Power and Human Reason: From Judgement to Calculation*. Harmondsworth: Pelican.

Agradecimientos

Agradezco a los evaluadores anónimos. También agradezco la discusión de este artículo a Felipe Morales Carbonell, Felipe Álvarez, Felipe Tobar y María Soledad Krause. Esta investigación fue financiada por el Proyecto ANID FONDECYT 1230128 Desconfianza: un Factor Causal de las Crisis Institucionales Searleanas.

Declaración de Autoría - Taxonomía CRediT	
Autor/es	**Contribuciones**
Rodrigo Alfonso González Fernández	Al tratarse de autoría única, la contribución total corresponde al mismo autor. El contenido presentado en el artículo es de exclusiva responsabilidad del autor.

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.

Declaración de uso de inteligencia artificial
Rodrigo Alfonso González Fernández **DECLARA** que la elaboración del artículo titulado "La factibilidad del niño-máquina de Turing" no contó con el apoyo de inteligencia artificial (IA).

Fecha de recepción: 14 de julio de 2021
Fecha de revisión: 15 de septiembre de 2021
Fecha de aprobación: 20 de abril de 2025
Fecha de publicación: 15 de julio de 2025

Sophia 39: 2025.

ISSN impreso:1390-3861 / ISSN electrónico: 1390-8626, pp. 115-139.